2014

Human face recognition under degraded conditions

Soodeh Nikan
University of Windsor

Human Face Recognition under Degraded Conditions

By

Soodeh Nikan

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctorate of Philosophy
at the University of Windsor

Windsor, Ontario, Canada

2014

Human Face Recognition under Degraded Conditions

by

Soodeh Nikan

APPROVED BY:

_____

Y. L. Murphey, External Examiner
University of Michigan-Dearborn

_____

B. Boufama
School of Computer Science

_____

M. Khalid
Department of Electrical and Computer Engineering

_____

J. Wu
Department of Electrical and Computer Engineering

_____

M. Ahmadi, Advisor
Department of Electrical and Computer Engineering

September 16, 2014

Declaration of Originality

DECLARATION OF CO-AUTHORSHIP

I hereby declare that this dissertation incorporates material that is the result of research conducted under the supervision of my supervisor, Dr. M. Ahmadi. Results related to this research are reported in Chapters 2 through 8.

I am aware of the University of Windsor's Senate Policy on Authorship and I certify that I have properly acknowledged the contributions of the other researchers to my dissertation, and I have obtained written permission from my co-author to include the aforementioned materials in my dissertation.

I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

DECLARATION OF PREVIOUS PUBLICATION

I declare that this dissertation includes seven original papers that have been previously published/ submitted for publication in peer reviewed conferences and journals as follows:

Dissertation Chapter	Publication Title/full Citation	Publication Status
Chapter 2	S. Nikan, M. Ahmadi, "Study of the effectiveness of various feature extractors for human face recognition for low resolution images," in *International Conference on Artificial Intelligence and Software Engineering (AISE'14)*, Phuket, Jan 2014, pp. 1-6.	Published
Chapter 3	S. Nikan, M. Ahmadi, "Effectiveness of various classification techniques on human face recognition," in *Proc. High Performance Computing and Simulation (HPCS), 2014 International Conference on*. pp. 651-655. IEEE, Jul 2014.	Published
Chapter 4	S. Nikan, M. Ahmadi, "Performance Evaluation of Different Feature Extractors and Classifiers for Recognition of Human Faces with Low Resolution Image," in *Proc. International Conference on Advanced Technology & Sciences (ICAT'14)*, Aug 2014, pp. 13-18.	Published
Chapter 5	S. Nikan, M. Ahmadi, "Recognition of human faces under different degradation conditions," Pan Stanford Publishing, Book Chapter, 24 pp, Apr 2014.	Accepted for publication
Chapter 6	S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 1699-1702. IEEE, Nov 2012.	Published
Chapter 7	S. Nikan, M. Ahmadi, "A local gradient-based illumination invariant face recognition using LPQ and multi-resolution LBP fusion," *IET Image Processing*, 10 pp, 2014, in press.	Published
Chapter 8	S. Nikan, M. Ahmadi, "Fusion of global and local based techniques for face recognition algorithm under degraded conditions," submitted to *Image and Vision Computing*, 34 pp, 2014.	Submitted

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my dissertation. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my dissertation.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

Comparative studies on the state of the art feature extraction and classification techniques for human face recognition under low resolution problem, are proposed in this work. Also, the effect of applying resolution enhancement, using interpolation techniques, is evaluated. A gradient-based illumination insensitive preprocessing technique is proposed using the ratio between the gradient magnitude and the current intensity level of image which is insensitive against severe level of lighting effect. Also, a combination of multi-scale Weber analysis and enhanced DD-DT-CWT is demonstrated to have a noticeable stability versus illumination variation. Moreover, utilization of the illumination insensitive image descriptors on the preprocessed image leads to further robustness against lighting effect. The proposed block-based face analysis decreases the effect of occlusion by devoting different weights to the image subblocks, according to their discrimination power, in the score or decision level fusion. In addition, a hierarchical structure of global and block-based techniques is proposed to improve the recognition accuracy when different image degraded conditions occur. Complementary performance of global and local techniques leads to considerable improvement in the face recognition accuracy. Effectiveness of the proposed algorithms are evaluated on Extended Yale B, AR, CMU Multi-PIE, LFW, FERET and FRGC databases with large number of images under different degradation conditions. The experimental results show an improved performance under poor illumination, facial expression and, occluded images.

Dedication

To my Dear Parents

Acknowledgement

There are several people that I owe my sincere appreciation for their generous help, guidance and contributions to completion of this doctoral dissertation.

First of all, I would like to express my gratitude to my advisor Dr. Majid Ahmadi for his invaluable advisedly support, inspiration and commitment throughout the course of this work.

I am also grateful to the rest of my doctoral committee: Dr. Jonathan Wu and Dr. Mohammed Khalid from the electrical and computer engineering department, Dr. Boubakeur Boufama from the school of computer science, and Dr. Yi Lu Murphey for their time and attention in reviewing my dissertation, and their encouraging words and constructive criticism.

Finally, my deepest gratitude goes to my loving husband, mom, dad, sister and brothers. Without their generous love, support, and encouragement, this research work would have never come to a successful completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In biological identification technologies physiological characteristics of individuals, such as face, finger print, retina, palm print and iris, are employed to distinguish their identity. Face images, compared to other biometrics, are conveniently collectable and human face recognition is a non-contact and convenient identification process. It includes a wide range of applications such as access control, social networking, security surveillance and border-crossing monitoring. Following by detection and localization of human faces in an image by segmenting the scene into face and non-face regions using shape information and template matching [1], the identity of an unknown individual is investigated by automatically processing their face image and comparing to the existing image database of labeled subjects to find the most similar individual. In law enforcement and forensic investigation applications it is crucial to recognize faces accurately and fast enough. However, the accuracy of identification is significantly reduced when the image is affected by degrading conditions in unconstraint situations, such as lighting effect, occlusion on the face, facial expressions, low resolution and blur, head pose variation and appearance changes due to aging effect. Therefore, this is a challenging research area to explore image processing and machine learning strategies to reduce degrading effects and increase the accuracy of face recognition.

Face recognition algorithm consists of preprocessing, feature extraction and classification. A preprocessing technique is utilized by employing image processing approaches to normalize the

image appearance and reduce degrading effects. Feature extraction describes the preprocessed image and explores its distinctive characteristics to reduce the processing data scale and complexity from image scale to feature space size. Image descriptors are divided into two categories, subspace learning approaches and local based feature extractors. Principal component analysis (PCA) [2], linear discriminant analysis (LDA) [3], locally sensitive discriminant analysis (LSDA) [4] and neighborhood preserving embedding (NPE) [5] are some examples of subspace learning based feature extractors. Local based descriptors derive distinctive local characteristics of image such as Gabor filtering [6], local binary pattern (LBP) [7] and local phase quantization (LPQ) [8]. Finally, a classification technique is utilized to compare the extracted features of the unknown individual image with the labeled images in the database and attributes the identity of the most similar subject to the image of unknown individual. Nearest neighbor (NN) [9, 10] classifier is a straight forward classifier where the distances between the unlabeled sample and all gallery samples with known identity labels are calculated using a distance metric such as Manhattan, which calculates the l_1 norm of the line length between two samples, Euclidean, where the l_2 norm of the distance is calculated, cosine and correlation, which determine the cosine of the angle and correlation between samples. Also, there exist some classification techniques based on the learning strategies such as sparse classifier (SC) [11], multi-layer feedforward neural network (MLFNN) [12] and extreme learning machine (ELM) [13].

Lighting direction can significantly affect the appearance of the face. The diversity between image samples of one class of identity under different illumination conditions can be larger than inherent differences between two subjects with distinct identities [14]. Thus, this degrading effect can decrease the accuracy of face recognition enormously. In order to deal with the illumination variation effect, several strategies have been employed in the literature. Illumination insensitive image preprocessing utilizes transformations at the intensity level or gradient based methods to remove the influence of lighting variation and enhance the face appearance [15]. An illumination insensitive image representation is proposed in this work based on the ratio of gradient magnitude to the image intensity level [16]. Also, we proposed the multi-scale Weber analysis which is the ratio of differences between each pixel and its neighbors at different neighborhood sizes over the current intensity level of pixel [17].

Another category of illumination insensitive preprocessing techniques are based on the image decomposition into frequency subbands where the luminance and reflection components are separable. In the proposed technique in this work, double density dual tree complex wavelet transform (DD-DT-CWT) as a powerful shift invariant and directionally selective function is employed in order to locally analyze image in multiresolution manner in time and frequency domain simultaneously. In the proposed method, we have applied DD-DT-CWT on the Weber normalized image in order to further reduce the effect of illumination variation, by enhancement in the contrast of low frequency subbands and enlarging image details of high frequency coefficients [17].

Illumination insensitive feature extractors such as LBP, LPQ and Gabor which extract characteristics of image, insensitive to lighting condition, can also be utilized to cope with the illumination variation. Nevertheless, compound approaches lead to superior performance by employing fusion of different methods similar to our proposed techniques in this work [16, 17].

In realistic applications, the performance of face recognition techniques are degraded due to the low resolution problem. However, capturing, saving and analysis of high resolution images increases the cost and computational complexity of the recognition system. The state of the art feature extraction and classification techniques are affected distinctively by resolution reduction. As illustrated in the proposed study on the performance of various techniques in this work [18-20] Gabor is a powerful feature extractor to extract a huge number of distinctive features. A dimensionality reduction technique is required to reduce the size of feature space. The performance of Fisher's linear discriminant (FLD) [21] improves by reducing the image size. However, discrete cosine transform (DCT) [22] and PCA show considerable robustness against resolution changes. The accuracy of Gabor technique is reduced by employing low resolution images. However, resolution enhancement using the proposed bilinear, bicubic and nearest neighbor interpolation methods [23] can improve the recognition accuracy to some extent [20]. In the proposed evaluation in this work [19, 20], ELM and SC show better performance under degrading conditions.

Occlusion is a challenging effect which is arisen in real-world situations where images are corrupted by irrelevant obstacles or objects [24]. Therefore, distinctive facial information are occluded. When, non-occluded parts of face do not provide sufficient discriminative data, the

identification process does not perform accurately [25]. In global-based face recognition techniques the whole face area is analyzed and the occluded parts affect the whole extracted information significantly [26]. While, block-based techniques which analyze face image locally by image partitioning, reduce the influence of degraded image sub regions on the identification process [27].

In the proposed technique in this work [28] the face image is partitioned into equal-sized and non-overlapping sub blocks. Uniform LBP descriptor is applied on the image sub-blocks to extract distinctive texture features from those areas separately. Chi-Square is employed as histogram similarity metric in local classifiers corresponding to different image blocks. Finally, a weighted majority voting scheme is used for decision fusion. Local entropy is proposed to devote weights to classifiers results according to the block informative richness. This way, we can reduce the effect of blocks with appearance deformation on the final decision. Experimental results show the significantly high recognition accuracy of our method on the challenging AR face database [29] compared to recent well-known approaches, without imposing computational complexity.

Similarly, image partitioning is applied on the preprocessed image in the proposed technique in this work [16] and LPQ and multi-scale LBP, extract the sub regions characteristics. Distance measurements of local nearest neighbor classifiers are fused at the score level to find the best match and decision-level fusion combines the results of two matching techniques. Entropy, class posterior probability and mutual information are utilized as the weights of fusion components. Simulation results on the Extended Yale B [30], AR [29], Multi-PIE [31] and FRGC [32] databases show the improved performance of the proposed algorithm under severe illumination and occlusion with low computational complexity and no reconstruction or training requirement.

Nevertheless, regarding the psychological point of view about the human perception of people identities based on their face image, fusion of global and local information leads to improved recognition accuracy [33, 34]. Robustness and accuracy against image degradation conditions is achieved using the proposed fusion of global-based and local-based face recognition strategies in the final part of this work [17]. In the global technique, Gabor filter and PCA are applied on the face image to extract features and for the classification, ELM was used. Furthermore, through the local-based approach, illumination insensitive LPQ is applied on the preprocessed image and subsequently the resultant image is divided into sub blocks. Local histograms of sub blocks are

fed into the sparse classifiers and local results are combined using a weighted majority voting decision fusion. The final class label is decided by finding the greater fuzzy membership value between the block-based and global-based results. Experimental results show the significant improvement of the proposed technique.

The chapters are organized as follows. In chapter 2 the effectiveness of various feature extractors for human face recognition for low resolution images is discussed. Chapter 3 describes the effectiveness of various classification techniques on human face recognition. Performance of different feature extractors and classifiers for recognition of human faces with low resolution and interpolated images is evaluated in chapter 4. In chapter 5 recognition of human faces under different degradation conditions is described. Chapter 6 provides human face recognition under occlusion using LBP and entropy weighted voting. A local gradient-based illumination invariant face recognition using LPQ and multi-resolution LBP fusion is determined in chapter 7. Chapter 8 explains the fusion of global and local based techniques for face recognition under degraded conditions. Finally, this work is concluded in chapter 9.

1.1 REFERENCES

[1] J. Wang, T. Tan, "A new face detection method based on shape information," *Pattern Recognition Letters*, vol. 21, pp. 463-471, 2000.

[2] M. Turk, A. Pentland, "Eigenfaces for recognition," *J. Cog. Neurosci.* vol. 3, pp. 71-86, 1991.

[3] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A.*, vol. 14, (8) pp. 1724-1733, 1997.

[4] D. Cai, X. He, K. Zhou, J. Han and H. Bao "Locality sensitive discriminant analysis," in *Proc. IJCAI'07*, 2007, pp. 708-713.

[5] X. He, D. Cai, S. Yan, H. J. Zhang, "Neighborhood Preserving Embedding," in *Proc. ICCV'05*, 2005, pp. 1208-1213.

[6] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression*," IEEE Trans. Acoust. Speech. Sign. Process.* Vol. 36, pp. 1169-1179, 1988.

[7] T. Ahonen, A. Hadid, M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 28, pp. 2037-2041,

December 2006.

[8] D. Zang, "Illumination invariant object tracking based on multiscale phase," in *Proc. IEEE International Conference on Image Processing (ICIP10)*, Hong Kong, 2010, pp. 357-360.

[9] H. Wang, "Nearest neighbors by neighborhood counting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 942-953, June 2006.

[10] M. H. Ramappa, S. Krishnamurthy, "Comparative study of different feature extraction and classification methods for recognition of handwritten Kannada numerals," *IJDTA.*, vol. 6, pp. 71-90, August 2013.

[11] J. Wright, A.Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 210-227, February 2009.

[12] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Great Britain: Springer, 2002.

[13] G. B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, pp. 513-529, April 2012.

[14] Y. Adini, Y. Moses, S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 19, pp. 721-732, 1997.

[15] H. Han, S. Shan, X. Chen, W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, (6) pp. 1691-1699, 2013.

[16] S. Nikan, M. Ahmadi, "A local gradient-based illumination invariant face recognition using LPQ and multi-resolution LBP fusion," *IET Image Processing*, 10 pp, 2014, in press.

[17] S. Nikan, M. Ahmadi, "Fusion of global and local based techniques for face recognition algorithm under degraded conditions," submitted to *Image and Vis. Comp.*, 34 pp, 2014.

[18] S. Nikan, M. Ahmadi, "Study of the effectiveness of various feature extractors for human face recognition for low resolution images," in *International Conference on Artificial Intelligence and Software Engineering (AISE'14)*, Phuket, Jan 2014, pp. 1-6.

[19] S. Nikan, M. Ahmadi, "Effectiveness of various classification techniques on human face recognition," in *Proc. High Perf. Comp. and Sim. (HPCS)*, Jul 2014, pp. 651-655.

[20] S. Nikan, M. Ahmadi, "Performance Evaluation of Different Feature Extractors and Classifiers for Recognition of Human Faces with Low Resolution Image," in *Proc. International Conference on Advanced Technology & Sciences (ICAT'14)*, Aug 2014, pp. 13-18.

[21] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell*., vol. 19, (7) pp. 711-720, 1997.

[22] W. Chen, M.J. Er, S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst. Man. Cyb.*, vol. 36, (2) pp. 458-466, 2006.

[23] S. M. Metev, V. P. Veiko, Practical Image and Video Processing Using MATLAB, 1st ed., Hoboken, New Jersey: John Wiley & Sons, 2011.

[24] T. Charoenpong, "A survey of face occlusion detection for visual surveillance system," in *Proc. Int. Conf. Knowl. Smart Tech., KST*, 2010, pp. 24-25.

[25] M. Storer, M. Urschler, H. Bischof, "Occlusion detection for ICAO compliant facial photographs," in *Proc. IEEE. Comp. Soc. CVPR. Wksp. (CVPRW)*, 2010, pp. 122-129.

[26] M. Rui, A. Hadid, J. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Proc. IEEE Int. Conf. Autom. FGR. (AFGR)*, 2011, pp. 442-447.

[27] T. Y. Kim, K. M. Lee, S. U. Lee, C. H. Yim, "Occlusion Invariant Face Recognition Using Two-Dimensional PCA," *Comm. Com. Inf. Sc*., vol. 4, pp. 305-315, 2007.

[28] S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," in *International Conference on Pattern Recognition (ICPR12)*, Tsukuba, November 2012, pp. 1699-1702.

[29] A. Martinez, R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998. Available (online) at: < http:// www2.ece.ohio-state.edu/~aliex/ARdatabase.html >.

[30] K. C. Lee, J. Ho, D. Kriegman, "Acquiring linear sub spaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, (5) pp. 684-698, 2005. The Extended Yale B face database, Available (online) at: <http://vision.ucsd.edu /~leekc/ExtYaleDatabase/ExtYaleB.html >.

[31] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, (5) pp. 807-813, 2010.

[32] P. J. Phillips, P. Flynn, T. Scruggs, et al., "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, San Diego, 2005, pp. 947-954.

Chapter 2

Study of the Effectiveness of Various Feature Extractors for Human Face Recognition for Low Resolution Images

Face recognition is widely used in video surveillance, law enforcement, security and human-computer interaction. In real world situations like video surveillance, the performance of face recognition systems are degraded due to the low resolution problem since it is expensive to capture and save high quality images and is computationally cost effective. In this chapter, we study the effect of image resolution reduction on the performance of some of the state of the art feature extraction algorithms. These algorithms in the literature can be divided into two different categories. The first group of techniques are appearance-based subspace methods which are dimensionality reduction approaches such as principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. PCA performs well when there is a small number of samples per individual subject but suffers from the intra-subject appearance variations. LDA considers the discrimination between classes in projection of data set to the new subspace and minimizes the intra-class changes, while its performance degrades with the small number of sample images. To solve this problem PCA is first applied to reduce the dimensionality prior to LDA, which is called Fisher's linear discriminant (FLD) technique [3]. The second group of feature extraction

techniques includes approaches in which the unique characteristics of image are captured by using a frequency domain transformation such as discrete cosine transform (DCT) [4] or Gabor wavelet [5]. In these techniques some of the frequency domain coefficients are used as the image features in the recognition step. Gabor wavelet suffers from high dimensionality of spatial frequency features which can be solved by one of the dimensionality reduction subspace approaches mentioned above. In this chapter, all of the mentioned feature extraction techniques in two categories are applied on the face database with different image resolutions and the nearest neighbor classifier finds the best matches to study the influence of low resolution on the face recognition performance. The rest of chapter is organized as follows. In second and third sections, the feature extraction techniques and nearest neighbor classifier are explained in details. The forth section, shows experimental results and comparisons. The chapter is concluded in final section.

2.1 FEATURE EXTRACTION TECHNIQUES

2.1.1 PCA

PCA is a linear holistic-based technique which is applied on the whole face image of size $R \times C$ and transforms the image data to the subspace with lower dimension which is called eigenspace. We first find the mean image of the training dataset and subtract the mean image from each individual image and find the covariance matrix as follows [1],

$$Cov = \frac{1}{N_I}\sum_{k=1}^{N_I}(I_k - M)\,.\,(I_k - M)^T, \tag{2.1}$$

where N_I is number of training images, M is the mean image and I_k is the k^{th} training sample.

The eigenvalues of the covariance matrix are sorted in decent order. The eigenvectors related to a few number of largest eigenvalues are used to transform the test and train data to the lower dimension subspace [1],

$$\lambda.V = Cov.V \quad , \quad \widehat{I_i} = V^T.(I_i - M), \tag{2.2}$$

where λ and V are the eigenvalue and eigenvector, respectively. I_i is the i^{th} train or test sample and $\widehat{I_i}$ is its projected eigenface.

2.1.2 FLD

Fisher linear discriminant approach is a linear technique which is a combination of PCA and LDA. In LDA method we aim at reducing the dimensionality and maximization of the ratio of between-class scatter to the within-class variation matrix as follows [2],

$$S_b = \sum_{k=1}^{N_c} N_{c_k}.(M_k - M).(M_k - M)^T \quad , \quad S_w = \sum_{k=1}^{N_c} \sum_{I_i \in C_i} (I_i - M_k).(I_i - M_k)^T , \tag{2.3}$$

$$V_{opt} = argmax\left(\frac{V^T.S_b.V}{V^T.S_w.V}\right), \tag{2.4}$$

where S_w and S_b are the within-class and between-class scatter matrices, respectively. N_c is the number of individuals, N_{c_k} is the number of samples in the k^{th} class and C_i is the i^{th} class. M and M_k are the mean images of the training set and the k^{th} class, respectively.

Then, the eigenvector V can be found as $S_b.V = \lambda.S_w.V$ [2].

However, since the rank of V is $N_I - N_c$ and much smaller than the number of pixels in the image ($R \times C$), PCA is applied, to reduce the dimension first to $N_I - N_c$, prior to LDA, then LDA technique projects data to dimension $N_c - 1$ [3].

2.1.3 Gabor

Gabor wavelet is a spatial frequency localization technique. Moreover, the shape of human visual cortex can be approximated by a 2-D Gabor filter and this filter gives the time information of the frequency components of signal. Gabor filter is a Gaussian modulated by a sinusoid as follows [5, 6],

$$\psi_{s,o}(x,y) = \frac{{f_s}^2}{\pi\gamma\mu} . e^{-\left[\left(\frac{{f_s}^2\hat{x}^2}{\gamma^2}\right)+\left(\frac{{f_s}^2\hat{y}^2}{\mu^2}\right)\right]} . e^{-(j2\pi.f_s.\hat{x})} , \quad \begin{cases} s = 0, 1, 2, \dots, S_{max} - 1 \\ o = 0, 1, 2, \dots, O_{max} - 1 \\ \hat{x} = +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \tag{2.5}$$

$\beta_o = (o\pi)/8$, $f_s = f_{max}/\left(\sqrt{2}\right)^s$ and f_{max} is the maximum frequency of the filter bank and γ and μ are the sharpness of x and y axis, respectively, and equal to $\sqrt{2}$. S_{max} and O_{max} are the maximum number of scale and orientation, respectively.

The convolution of image and Gabor filter gives $S_{max} \times O_{max}$features for each image pixel pixel. Although the Gabor feature extraction method is more robust against appearance variations, compare to subspace techniques, it suffers from high dimensionality problem which can be solved by a dimensionality reduction technique [7].

2.1.4 DCT

Discrete cosine transform is a redundancy reduction method based on a set of orthogonal cosine functions as follows [4],

$$D(p,q) = \alpha(p)\alpha(q)\sum_{m=0}^{R-1}\sum_{n=0}^{C-1} I(m,n)\cos\left[\frac{(2m+1)p\pi}{2R}\right]\cos\left[\frac{(2n+1)q\pi}{2C}\right],$$

$$\text{where} \quad \begin{cases}\alpha(p) = 1/\sqrt{R} & if\ p = 0.\\ \alpha(p) = \sqrt{2/R} & if\ p = 1,2,\dots,R.\end{cases} \quad \begin{cases}\alpha(q) = 1/\sqrt{C} & if\ q = 0.\\ \alpha(q) = \sqrt{2/C} & if\ q = 1,2,\dots,C.\end{cases} \tag{2.6}$$

Applying DCT transformation on an image I of size $R \times C$ gives a matrix of frequency coefficients with the same size as image. The important characteristics of image are compressed in the small number of coefficients on the top left of features matrix. Therefore, a feature selection technique is applied to truncate the frequency coefficients to reduce the feature vector dimensionality. Zigzag static is employed as illustrated in Fig 2.1. A small number of coefficients are chosen as the feature vector. The number of selected features can affect the face recognition performance and is chosen regarding the experimental results [4].

2.2 NEAREST-NEIGHBOR CLASSIFIER

The behavior of the feature extraction techniques can be examined by using the nearest neighbor classifier where a distance metric is applied to find the distances between the feature vectors of the test image and the training samples and find the best match with the minimum distance. Euclidean distance is used in this chapter as the distance measurement metric [4].

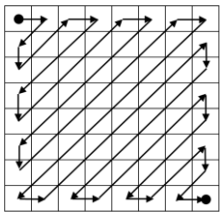

Figure 2.1. The zigzag manner which is applied to sort the components in DCT coefficient matrix.

2.3 EXPERIMENTAL RESULTS

In order to evaluate the performance of different face recognition techniques with variation in the image resolution, we use AT&T or Olivetti research lab (ORL) database which consists of 400 images of 40 people (10 per individual) of size 112×92 with different time, scales and illumination conditions, and various facial expressions (open/close eyes and smiling/not smiling) and up to 20 degree tilting and rotation [8, 9]. We use the first 5 samples of each person in the train set and the rest 5 images in the test dataset. Therefore, we have 200 images in the train set and 200 in the test set. The eigenspace dimension for PCA and FLD in all experiments is equal to 199 and 39, respectively. Also, according to the experimental results, the best DCT performance is achieved with 60 first features. Gabor filter has 8 scales, 5 orientations and $f_{max} = 0.25$. The image resolution is changed from 112×92 to 80×64, 40×32, 20×16 and 10×8 pixels. Some samples are illustrated in Fig 2.2. Table 2.1 and Fig 2.3 and 2.4 show the performance of all 5 techniques. As illustrated, the Gabor-FLD method has the best total performance, while, for both Gabor techniques, accuracy diminishes by reducing the image resolution. The performance of FLD improves by reducing the image size. It happens because appearance differences between samples of an individual decreases at small sizes and discrimination power of FLD increases [10]. However, DCT and PCA show considerable robustness against resolution changes.

Figure 2.2. Sample images of the ORL database at different resolutions.

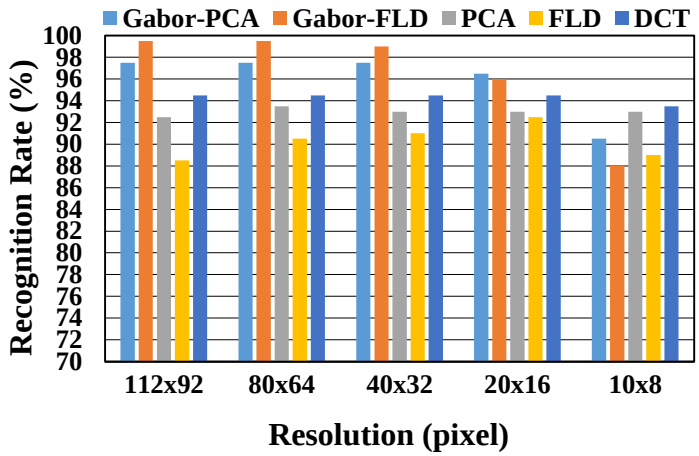

Figure 2.3. Recognition accuracy percentage of feature extraction methods versus image resolution.

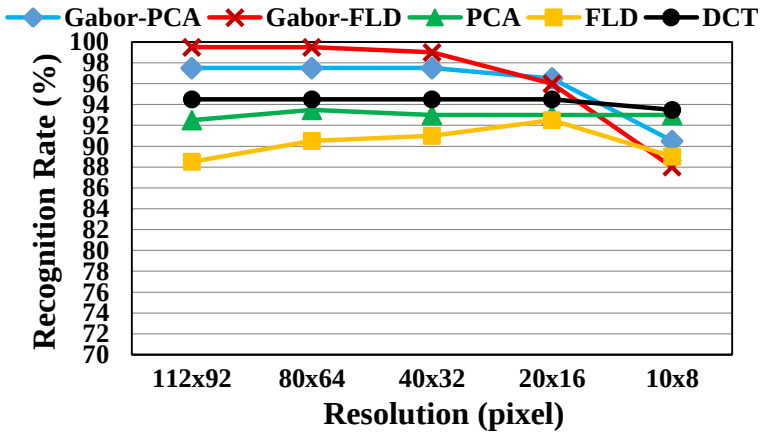

Figure 2.4. Recognition rates (%) of feature extraction techniques against various image resolutions.

Table 2.1. Identification error rates of feature extraction approaches for different image sizes.

Resolution [pixel] / Method	112x92	80x64	40x32	20x16	10x8
Gabor-PCA	0.025	0.025	0.025	0.035	0.095
Gabor-FLD	0.005	0.005	0.01	0.04	0.12
PCA	0.075	0.065	0.07	0.07	0.07
FLD	0.115	0.095	0.09	0.075	0.11
DCT	0.055	0.055	0.055	0.055	0.065

2.4 CONCLUSION

In this chapter, some of the well-known feature extraction techniques in the literature are examined at different image resolution, the issue which we may encounter in real world

applications. The ORL database with various image sizes is employed to investigate the effectiveness of DCT, PCA, FLD and Gabor techniques. The first two methods have almost robust performance at different resolutions. FLD has more discriminative power when reducing the image size. However, Gabor filter which benefits from PCA/FLD to overcome the high dimensionality problem, has superior performance in comparison with other techniques but it is sensitive to the image resolution and its accuracy is reduced by image size deduction.

2.5 REFERENCES

[1] M. Turk, A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neuro*. vol. 3, pp. 71-86, 1991.

[2] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*., vol. 14, (8) pp. 1724-1733, 1997.

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell*., vol. 19, (7) pp. 711-720, 1997.

[4] W. Chen, M.J. Er, S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst. Man. Cyb*., vol. 36, (2) pp. 458-466, 2006.

[5] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression," *IEEE Trans. Aco. Speech. Sig. Process*. Vol. 36, (7) pp. 1169-1179, 1988.

[6] L. Shen, L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Anal. Applic*., vol. 9, (2-3) pp. 273-292, 2006.

[7] V. Štruc, N. Pavešic, "The complete Gabor-Fisher classifier for robust face recognition," *EURASIP Advan. Signal. Process*., vol. 26, 2010.

[8] Available at: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

[9] T. Mandal, Q.M.J. Wu, Y. Yuan, "Curvelet based face recognition via dimension reduction," *Sig. Process*., vol. 89, (12) pp. 2345-2353, 2009.

[10] Y. Xu, Z. Jin, "Down-sampling face images and low-resolution face recognition," in *IEEE Int. Conf. Innov. Comp. Inf. and Control (ICICIC08)*, Dalian, China, 2008, pp. 392.

Chapter 3

Effectiveness of Various Classification Techniques on Human Face Recognition

Face recognition is an effective biometric identification technique with noticeable capability and effectiveness. Law enforcement, border security, video surveillance and human-computer interaction are some of its wide range of applications. In this chapter, we study the effect of various classification techniques on the performance of face recognition algorithm using fusion of some of the state of the art feature extraction methods. Gabor wavelet is a feature extraction technique in which the local distinctive characteristics of image are extracted by using the frequency domain transformation [1]. However, it suffers from the high dimensionality problem which can be solved by one of the dimensionality reduction subspace techniques such as principal component analysis (PCA) [2] or Fisher's linear discriminant (FLD) [3]. Local binary pattern (LBP) [4] is a well-known local texture descriptor which is easy to execute with simple calculations and invariant against monotonic image degradations, illumination and noise. Combination of this technique with Gabor leads to significant performance to extract more distinctive and robust features out of face image.

Classification method can affect the performance of face recognition. The classification techniques are divided into two main categories. The first group are nonparametric classifiers such as different forms of nearest neighbor method with various distance metrics and classifiers

based on the sparse representation which are the general form of nearest neighbor approach [5]. The second category of matching techniques includes parametric methods inspired by the biological human nervous systems [6]. Artificial neural network and extreme learning machine [7] are some examples in this regard.

In this chapter, all of the mentioned classification techniques in both categories are applied on the face databases with different illumination conditions, facial expression and large number of individuals. The rest of chapter is organized as follows. In the second and third sections, the feature extraction techniques and classification approaches are explained in details. The forth section, shows the experimental results and comparisons. The chapter is concluded in the final section.

3.1 FEATURE EXTRACTION

3.1.1 Gabor

Two dimensional Gabor wavelet is a spatial frequency localization filter which resembles the human visual cortex. 2-D Gabor filter extracts the time information of the frequency components of the image and can be approximated by a multiplication of a sinusoid and a Gaussian function as follows [1, 8].

$$\psi_{s,o}(x,y) = \frac{{f_s}^2}{\pi\gamma\mu} \cdot e^{-\left[\left(\frac{{f_s}^2\hat{x}^2}{\gamma^2}\right)+\left(\frac{{f_s}^2\hat{y}^2}{\mu^2}\right)\right]} \cdot e^{-(j2\pi.f_s.\hat{x})},$$

$$where \begin{cases} s = 0,1,2,\dots,S_{max}-1 \\ o = 0,1,2,\dots,O_{max}-1 \\ \hat{x} = +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \tag{3.1}$$

S_{max} and O_{max} are the maximum number of scale and orientation, respectively. $\beta_o = (o\pi)/8$ and $f_s = f_{max}/\left(\sqrt{2}\right)^s$, where f_{max} is the maximum frequency. The sharpness of x and y axis are defined by γ and μ, respectively, which are equal to $\sqrt{2}$ [9]. Gabor coefficients of the image are extracted by convolution of the image and a bank of $S_{max} \times O_{max}$ Gabor filters. Although the Gabor coefficients are robust against appearance variations, Gabor technique suffers

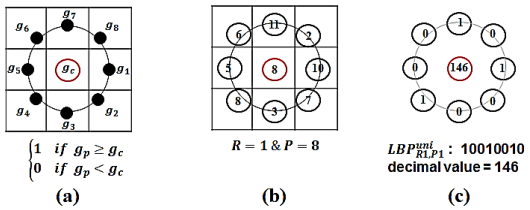

Figure 3.1. Different steps of local binary pattern: (a) neighborhood around each pixel, (b) example of 8 neighbors on the neighborhood with radius 1 and (c) binary pattern and decimal value of the centre pixel.

from high dimensionality problem [8, 9].We require a down sampling method by a factor which is found experimentally through the exhaustive search regarding the performance of the face recognition algorithm. Gabor images are resized using bicubic interpolation by down sampling factor along vertical and horizontal dimensions

Feature vector for each image is constructed by concatenation of Gabor coefficients.

Dimensionality reduction technique such as PCA and FLD, are also applied on the feature vector to reduce the computational cost, which are called Gabor PCA (GPCA) and Gabor FLD (GFLD), respectively [9].

3.1.2 Local Binary Pattern

LBP is one of the most important texture descriptors in local appearance based approaches, which is insensitive to illumination variation and noise effect since it compares each pixel with P pixels on a neighborhood around that with radius R. Therefore, it eliminates the pixel offsets [4]. If the neighbor is equal or greater than the center pixel it is replaced by 1 and if it is smaller, it is substituted by 0. Thus, we have a P-bit binary pattern for each center pixel which gives a decimal value, as illustrated in Fig 3.1. There is no training requirement and the procedure is applied directly on image pixels which makes it fast and easily integrable into the new data sets [4, 10].

3.1.3 Local Gabor Pattern

In order to benefit from the advantages of LBP descriptor and Gabor wavelet simultaneously, in this chapter we use the combination of these techniques to extract more precise and distinctive

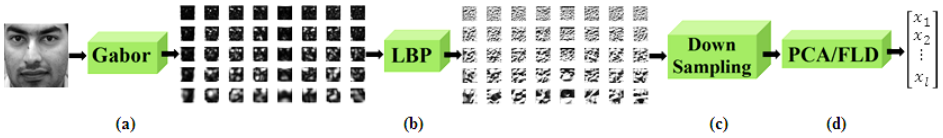

Figure 3.2. Proposed LGP technique to extract feature vector of image: (a) Gabor filtering of the image, (b) applying LBP on each filtered image, (c) down sampling of local Gabor patterns and (d) dimensionality reduction using PCA or FLD.

features out of face image [11]. We apply the local binary pattern technique on each Gabor-filtered image to reduce the illumination and noise effect. In order to reduce the feature dimensionality, we employ down sampling to resize the Gabor pattern images and concatenate all filtered images to construct the feature vector. Furthermore, PCA or FLD approaches are applied to reduce the large dimensionality of the feature vector. The feature extraction technique is illustrated in Fig 3.2.

3.2 CLASSIFICATION TECHNIQUES

3.2.1 Nearest-Neighbor Classifier

The extracted features of the gallery and probe images can be compared to find the best match in the classification stage. Nearest neighbor classifier is a simple nonparametric classifier where a geometrical distance metric is applied to find the distances between the feature vectors and find the nearest neighbor with the minimum distance measurement. Manhattan (City block), Euclidean, Cosine and Correlation cost functions are employed as distance measurement metrics [12, 13], as shown in Table 3.1.

3.2.2 Sparse Classifier

Sparse classifier is the general form of nearest neighbor classifier with considerable discrimination power. This technique is based on the sparsest representation of probe image of the unknown individual [5]. The probe sample can be represented as a sparse linear combination

Table 3.1. Various distance measurement metrics of the nearest neighbor classifier.

Metric	$\boldsymbol{D(G,P)}$
Manhattan(l_1-norm)	$\sum_{i=1}^{N_f} \lvert g_i - p_i \rvert$
Euclidean (l_2-Norm)	$\left[\sum_{i=1}^{N_f} (g_i - p_i)^2\right]^{1/2}$
Correlation	$\sum_{i=1}^{N_f} \frac{(g_i - \mu_g).(p_i - \mu_p)}{\sigma_g \sigma_p}$
Cosine	$\frac{\sum_{i=1}^{N_f} g_i.p_i}{\sqrt{\sum_{i=1}^{N_f} g_i^2 . \sum_{i=1}^{N_f} p_i^2}}$
G and P are gallery and probe images and g_i and p_i are their feature elements, respectively. N_f is the number of extracted features from those images. μ_g, σ_g, μ_p and σ_p are mean and standard deviation of gallery and probe images, respectively.	

of images with same identity in the gallery set. The solution to the l_1-norm minimization problem is utilized in order to find the sparsest representation [5]. The matrix of gallery images is as follows,

$$G = [G_1, G_2, G_3, \dots, G_N], \tag{3.2}$$

$$G_i = \left[F_1^g, F_2^g, F_3^g, \dots, F_M^g\right], \tag{3.3}$$

where G_i is the matrix of gallery images of the i^{th} class and F_k^g is the feature vector of the k^{th} gallery image in G_i. The number of classes and gallery samples per class are N and M, respectively.

A probe image can be represented as follows,

$$F_i^p = G.C, \tag{3.4}$$

$C = [0,0,\dots,0,c_1^i, c_2^i, \dots, c_M^i, 0,0,\dots,0]$, where $c_j^i, j \in [1,2,\dots,M]$, is the j^{th} coefficient associated with the i^{th}class.

The class label of the probe image is found as the sparsest solution of the l_1 -norm minimization problem as follows,

$$(l_1): \quad \hat{C}_1 = rgmin\|C\|_1, \ while \ F^p = G.C. \tag{3.5}$$

3.2.3 Multilayer Feed-Forward Neural Network (MFNN)

Artificial neural network models the intellectual human nerve system. It consists of input layer, where the number of neurons is equal to the length of feature vector, one or more hidden layer and an output layer, where the number of neurons is equal to the number of data classes. In this chapter we proposed a multilayer feed-forward network with one hidden layer and scaled conjugate gradient back propagation approach as the training function. It has a fast convergence for pattern recognition problems with large number of data classes since the search speed along the conjugate direction is higher than the steepest descent direction and it requires small memory [6]. The adequate number of neurons in the hidden layer is obtained experimentally through an exhaustive search.

3.2.4 Extreme Learning Machine (ELM)

Despite the slow learning and poor scalability of the feed forward neural networks due to the iterative weights tuning using slow gradient decent technique, extreme learning machine is very fast. ELM is a single hidden layer neural network with random number of hidden nodes with determined weights and no requirement to adjust the weights iteratively [7]. Thus, ELM is remarkably faster compared to feed forward neural networks with gradient decent learning approach, and speed is an important factor in many pattern recognition applications.

3.3 EXPERIMENTAL RESULTS

In order to evaluate the performance of feature extractor algorithms with various classification techniques, we use the Extended Yale B, FERET and Multi-PIE face databases with large number of images under variations in illumination, expression and image quality. Gabor filter has 8 scales, 5 orientations, $f_{max} = 0.25$ and down sampling factor is 25.

3.3.1 Extended Yale B Database

The Extended Yale B database [14] contains images of size 192×168 pixels for 38 subjects with 9 poses and 64 illumination conditions per pose. The frontal pose images are utilized in this work. Database is divided into five subsets corresponding to the angle between the light source and the camera axis, α. Subset 1 consists of 7 images per subject, ($\alpha < 12^{\circ}$). Subsets 2 to 5 contain 12, 12, 14 and 19 images per subject, with ($20^{\circ} < \alpha < 25^{\circ}$), ($35^{\circ} < \alpha < 50^{\circ}$), ($60^{\circ} < \alpha < 77^{\circ}$) and ($\alpha > 78^{\circ}$), respectively. Thus, subsets 1 to 5 contain 226, 454, 454, 532 and 722 faces, respectively. Figure 3.3a shows the images of the first individual in subset 5. Subset 1 is used as the gallery set and subsets 2-5, as probe sets [5]. Figure 3.4 shows the performance of feature extractors with different classification techniques.

3.3.2 FERET Database

Face Recognition Technology (FERET) program database [15] is a huge database consists of 14,051 grayscale images in different subsets based on various illumination, facial expression and pose conditions. In this experiment we use subset ba, bj and bk of 200 individuals and one image per subject in each subset. Subset ba includes regular frontal images. Subset bj consists of alternative frontal images corresponding to ba set. Subset bk also contains frontal images corresponding to ba but with different lighting conditions. 400 images in subsets bj and bk are used as the gallery set and subset ba is the probe set. Images are of size 128×128 pixels. Figure

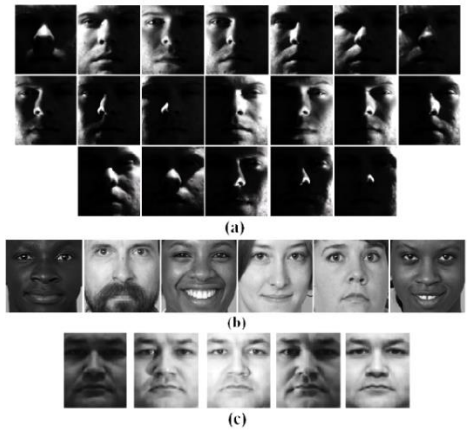

Figure 3.3. Sample images of face databases: (a) Extended Yale B subset 5, (b) FERET and (c) Multi-PIE.

Table 3.2. Recognition accuracy (%) of four feature descriptors with seven classification methods for FERET database.

Descriptor / Classification	GPCA	GFLD	LGP PCA	LGP FLD
NN (Manhattan)	95.5	93.5	94.5	86.5
NN (Euclidean)	89	92.5	91.5	87.5
NN (Cosine)	91.5	98.5	93	98
NN (Correlation)	91.5	98.5	93	98
Sparse Classifier	96	98.5	96.5	98
MFNN	99	98.5	98.5	98.5
ELM	99	98.5	99	98.5

3.3b shows some samples of the probe images. Table 3.2 illustrates the performance of feature extraction methods with different classification techniques.

3.3.3 Multi-PIE Database

In the CMU Multi-PIE database, there are 750,000 images of 337 individuals in four different sessions with 15 pose and 19 illumination conditions and various facial expressions [16]. Some samples of the first subject are shown in Fig 3.3c. All frontal pose images of 249 subjects with seven illumination conditions {0, 1, 7, 13, 14, 16 and 18} and neutral expression in session 1 were used as the gallery set. Four probe sets include ten frontal images of even-number illumination conditions from sessions 2-4 [17]. The gallery and probe sets include 1734, 2030, 2300 and 2390 images of size 128×128, respectively. Figure 3.5 shows the effect of different classification techniques on three probe sets.

3.4 CONCLUSION

In this chapter, four feature extraction methods, based on combination of Gabor wavelet, LBP and dimensionality reduction approaches, GPCA, GFLD, LGP-PCA and LGP-FLD, are examined with different classification techniques. The Extended Yale B, FERET and Multi-PIE databases with large number of subjects and various illumination and expression conditions are

employed to evaluate the effectiveness of various classifiers. The nearest neighbor classifier shows better results with cosine and correlation metrics in medium illumination variation and with Euclidean in severe lighting variation. Sparse classifier outperforms nearest neighbor with l_1 and l_2 norm cost functions, in most cases. ELM and MFNN have more significant recognition accuracy in all experiments.

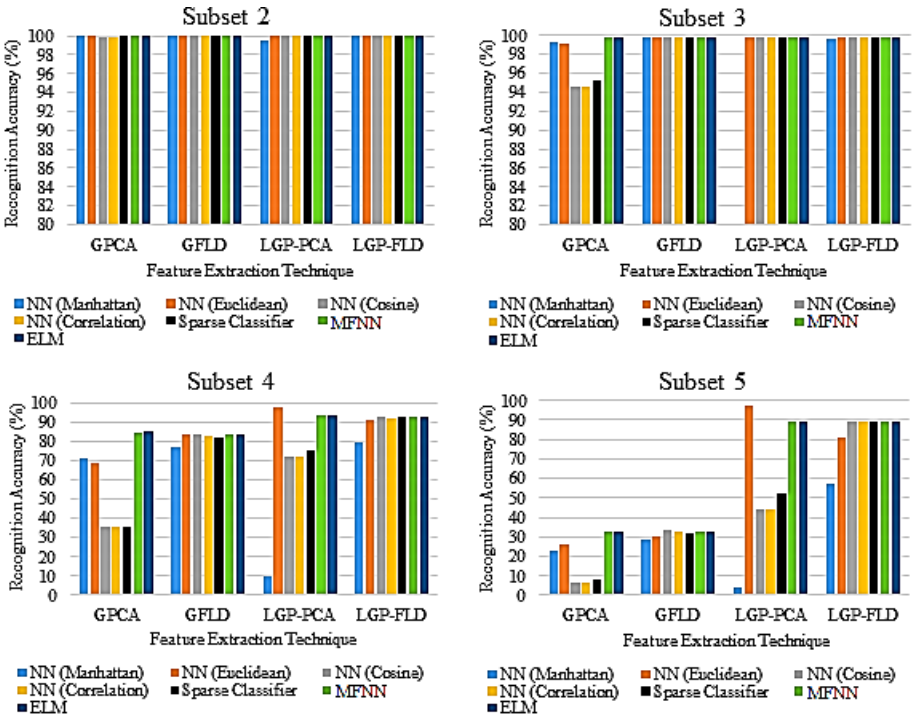

Figure 3.4. Identification accuracy (%) of four feature extraction techniques with seven classification methods for Extended Yale B database subsets.

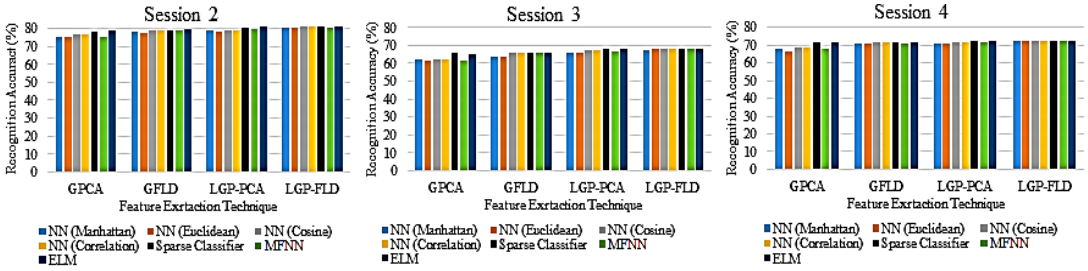

Figure 3.5. Recognition accuracy (%) of four feature extraction approaches with seven classification methods for Multi-PIE database subsets.

3.5 REFERENCES

[1] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, pp. 1169-1179, July 1988.

[2] M. Turk, A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci*., vol. 3, pp. 71-86, 1991.

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. sherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 19, pp. 711-720, July 1997.

[4] T. Ahonen, A. Hadid, M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 28, pp. 2037-2041, December 2006.

[5] J. Wright, A.Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, pp. 210-227, February 2009.

[6] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Great Britain: Springer, 2002.

[7] G. B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern., Syst*., vol. 45, pp. 513-529, April 2012.

[8] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Anal. Applic*., vol. 9, pp. 273-292, October 2006.

[9] S. Nikan, M. Ahmadi, "Study of the effectiveness of various feature extractors for human face recognition for low resolution images," in *International Conference on Artificial Intelligence and Software Engineering (AISE'14)*, Phuket, Jan 2014, pp. 1-6.

[10] S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," in *International Conf. on Pattern Recognition (ICPR12)*, Tsukuba, November 2012, pp. 1699-1702.

[11] Z. Lei, S. Liao, M. Pietikäinen, S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Trans. Image Process*., vol. 20, pp. 247-256, January 2011.

[12] H. Wang, "Nearest neighbors by neighborhood counting," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 28, pp. 942-953, June 2006.

[13] M. H. Ramappa, S. Krishnamurthy, "Comparative study of different feature extraction and classification methods for recognition of handwritten Kannada numerals," *IJDTA.*, vol. 6, pp. 71-90, August 2013.

[14] K.C. Lee, J. Ho, D. Kriegman, "Acquiring linear sub spaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 27, pp. 684-698, May 2005. The Extended YaleB face database, Available (online) at:

<http://vision.ucsd.edu /~leekc/ExtYaleDatabase/ExtYaleB.html >.

[15] P.J. Phillips, H. Moon, P.J. Rauss, S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell*., 22, pp. 1090–1104, October 2000.

[16] R. Gross, I. Matthews, J. Cohn, T. Kanade and S. Baker, "Multi-PIE," *Image Vision Comput*., vol. 28, pp., 807-813, May 2010.

[17] M. Yang, L. Zhang, S. C. Shiu and D. Zhang, "Robust kernel representation with statistical local features for face recognition," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 24, pp. 900-912, June 2013.

Chapter 4

Performance Evaluation of Different Feature Extractors and Classifiers for Recognition of Human Faces with Low Resolution Images

Among the biometric identification technologies, face recognition is the most popular one due to its remarkable effectiveness and broad applications in video surveillance, law enforcement and human-computer interaction. The image degradations caused by illumination, pose and aging as well as variations in uncontrolled environment, affect the accuracy of face recognition algorithms significantly. Low-resolution images which are captured by video surveillance cameras, lead to noticeable decline in the identification performance. However, handling high resolution images and their storage is not cost effective and can increase the computational complexity of the recognition algorithm. In this chapter, we examine the performance of different feature extraction techniques in combination with variety of classifiers with high and low resolution images. Furthermore, we evaluate the effect of image enhancement using various interpolation methods to improve the recognition accuracy. 2-D Gabor is a global feature descriptor where a Gaussian modulated by a sinusoid extracts the spatial frequency details of image [1]. Although it is a powerful descriptor to capture illumination invariant features, the high dimensionality of the feature vector burden computational complexity.

To reduce the size of feature vector, we use dimensionality reduction techniques, such as principal component analysis (PCA) [2], linear discriminant analysis (LDA) [3], locally sensitive discriminant analysis (LSDA) [4] and neighbourhood preserving embedding (NPE) [5]. We can also benefit from the local characteristic of local phase quantization (LPQ) [6] and combine it with Gabor filter to reduce the effect of blur variation.

In the classification stage we examine the performance of four different classifiers. Regularized discriminant classifier (RDC) [7], fuzzy nearest neighbour (FNN) [8], sparse classifier (SC) [9] and extreme learning machine (ELM) [10] are utilized in this chapter. Low resolution images are enhanced using the interpolation techniques such as nearest neighbour, bilinear and bicubic interpolation [11].

The rest of chapter is organized as follows. In section 4.1 and 4.2, feature extraction approaches are explained. Different classifier techniques are explained in section 4.3. Section 4.4 is related to the interpolation methods to enhance the image resolution. Section 4.5, shows the experimental results and comparisons. The chapter is concluded in section 4.6.

4.1 FEATURE EXTRACTION

4.1.1 Gabor

Representation of 2-D Gabor resembles the human visual system and a set of Gabor filters can extract image characteristics in different orientations and frequencies which are robust against appearance variations. Gabor filter is a Gaussian kernel multiplied by a sinusoid function as follows [1, 12],

$$\psi_{s,o}(x,y) = \frac{{f_s}^2}{\pi\gamma\mu} \cdot e^{-\left[\left(\frac{{f_s}^2\hat{x}^2}{\gamma^2}\right)+\left(\frac{{f_s}^2\hat{y}^2}{\mu^2}\right)\right]} \cdot e^{-(j2\pi.f_s.\hat{x})}, \tag{4.1}$$

where

$$\begin{cases} s = 0, 1, 2, \dots, S_{max} - 1 \\ o = 0, 1, 2, \dots, O_{max} - 1 \\ \hat{x} = +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \tag{4.2}$$

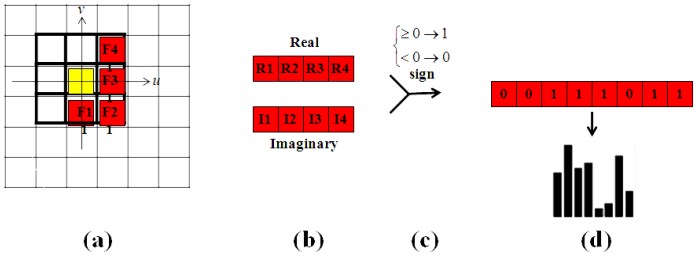

Figure 4.1. LPQ procedure: (a) STFT at four specific frequencies inside a small window at the pixel position, (b) finding the real and imaginary of the four STFT values, (c) binary quantization based on the sign of real and imaginary values and (d) constructing a histogram based on the 8-bit stream.

$\beta_o = (o\pi)/8$, $f_s = f_{max}/(\sqrt{2})^s$ where, f_{max} is the maximum frequency of the filter. S_{max} and O_{max} are the maximum number of scale and orientation, respectively. γ and μ, both equal to $\sqrt{2}$, are the sharpness of x and y axis, respectively. Therefore, we have a set of $S_{max} \times O_{max}$ filters which are convolved with the image to obtain the Gabor images. All Gabor coefficients are combined as a high dimensional feature vector. To reduce the size of feature vector and computation cost, a dimensionality reduction approach is applied such as PCA, LDA, LSDA and NPE [12, 13].

4.1.2 GLPQH

Combination of Gabor filter with LPQ leads to distinctive local image characteristics which are blur invariant. LPQ, which is applied on the Gabor images, is a well-known histogram-based and blur invariant feature extractor that performs the local phase analysis at four frequency components in a small neighbourhood around each image pixel [6]. The point spread function (PSF) of blur effect is centrally symmetric and its phase is zero at small frequencies within the bandwidth of PSF and thus the phase analysis is blur invariant. Therefore, assessment and binary quantization of the real and imaginary parts of the frequency components of short time Fourier transform (STFT) components at those frequencies, gives an 8-bit pattern for each pixel of image [6] as shown in Fig 4.1. Thus, we have a histogram of 256 bins for each Gabor image, concatenation of which leads to a feature vector of size $S_{max} \times O_{max} \times 256$. Dimensionality reduction techniques in section 4.2 are employed to reduce the size of feature vector. The whole procedure is shown in Fig 4.2.

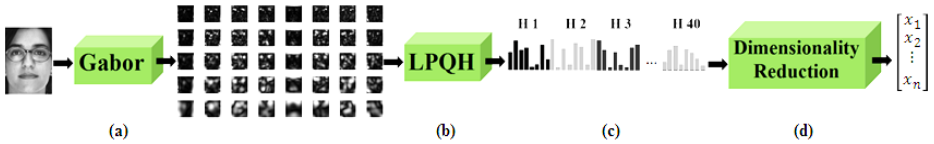

Figure 4.2. Proposed GLPQH technique to extract feature vector of image: (a) Gabor filtering of the image, (b) applying LPQ on each filtered image and find the histograms, (c) concatenate the histograms and (d) dimensionality reduction PCA, LDA, LSDA or NPE.

4.2 DIMENSIONALITY REDUCTION TECHNIQUES

4.2.1 Principal Component Analysis (PCA)

PCA is an unsupervised linear dimensionality reduction approach. This holistic-based technique utilizes the whole face image and transforms the image pixels to eigenspace which is a lower dimension subspace [2]. The mean image is calculated by averaging the gallery set. After subtracting the mean image from the gallery samples, the covariance matrix is calculated as follows [2, 12],

$$Cov = \frac{1}{N_I}\sum_{k=1}^{N_I}(I_k - M)\,.\,(I_k - M)^T, \tag{4.3}$$

where I_k is the k^{th}gallery sample, N_I is the number of gallery images and M is the mean image. The eigenvalues are sorted in descent order. The eigenvectors of the covariance matrix corresponding to a few number of largest eigenvalues are utilized to transform the gallery and probe images to eigenspace, which are called eigenfaces [2, 12],

$$\lambda.V = Cov.V \quad , \quad \widehat{I_i} = V^T.(I_i - M), \tag{4.4}$$

where λ and V are the eigenvalue and eigenvector, respectively. I_i is the i^{th} image sample and $\widehat{I_i}$ is its corresponding eigenface [12].

4.2.2 Linear Discriminant Analysis (LDA)

LDA reduces dimension linearly by maximizing the ratio of between-class to within-class scatter

matrix as follows [3],

$$S_b = \sum_{k=1}^{N_c} N_{c_k}.(M_k - M).(M_k - M)^T, \quad (4.5)$$

$$S_w = \sum_{k=1}^{N_c} \sum_{I_i \epsilon C_i} (I_i - M_k).(I_i - M_k)^T, \quad (4.6)$$

$$V_{opt} = argmax\left(\frac{V^T.S_b.V}{V^T.S_w.V}\right), \quad (4.7)$$

where N_c and $N_{c_k}$are the number of classes and samples in the k^{th} class, respectively. C_i is i^{th} subject, M and M_k are the mean of gallery samples and samples in k^{th} class, and S_w and S_b are the within and between-class scatter matrices respectively. $S_b.V = \lambda.S_w.V$ gives the eigenvector, V, [3, 12].

4.2.3 Locally Sensitive Discriminant Analysis (LSDA)

LSDA is a linear data analysis technique to reduce the data dimension, with small number of gallery samples, based on the local geometrical structure of data according to the adjacency graph where the data points are nodes and there are edges with weights equal to one between each node and its k nearest neighbours [4]. Due to the fact that in this method, the emphasis is on data discrimination, the graph of neighbours with the same class label is separated from the between class adjacent subset [4]. The aim is to find a data projection solution to map the adjacent graphs in such a way so that within-class nodes be as close as possible and between-class nodes stay as far as possible. Thus, the following maximum eigenvalue problem is required to be solved and the projection matrix is constructed of the eigenvectors, v, corresponding to m highest eigenvalues, λ, which are sorted in order [4],

$$S(\alpha\Phi_b + (1 - \alpha W_w))S^T v = \lambda S D_w S^T v, \quad (4.8)$$

where $S = (s_1, s_2, \dots, s_N)$ is the set of sample points, W_w and D_w are the within-class graph weight matrix and its column-sum, respectively. $\Phi_b = W_b - D_b$, where, W_b and D_b are the

between-class graph weight matrix and its column-sum, respectively, and $V_{N\times m}$ is the projection matrix (mathematical details can be found in [4]).

4.2.4 Neighbourhood Preserving Embedding (NPE)

NPE is a linear dimensionality reduction method where we first construct the adjacency graphs with N sample points as the nodes. There would be edges from each node to its k nearest neighbours. The weights on adjacent edges are calculated by minimization of the objective function as follows [5],

$$\psi(w) = min \sum_i \left\| s_i - \sum_j w_{ij} s_j \right\|^2. \tag{4.9}$$

Then we need to find the projection matrix by calculating eigenvalues λ, and eigenvectors v,

$$SLS^T v = \lambda SS^T v, \tag{4.10}$$

where $S = (s_1, s_2, \dots, s_N)$ is the set of sample points, $L = (I - w)^T (I - w)$ and I is the Identity matrix. Thus, the projected sample is calculated using the projection matrix V of size $N \times m$, which is composed of the eigenvectors corresponding to m largest eigenvalues [5],

$$\widehat{s_i} = V^T . s_i \,, \qquad where \quad V = \ (v_1, v_2, \dots, v_m). \tag{4.11}$$

4.3 CLASSIFICATION TECHNIQUES

4.3.1 Regularized Discriminant Classification (RDC)

Discriminant analysis classification is based on the normal distribution as follows [7],

$$G_i(\bar{F}) = (2\pi)^{-\frac{N}{2}} |\delta_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\bar{F}-\mu_i)^T \delta_i^{-1} (\bar{F}-\mu_i)}, \tag{4.12}$$

where μ_k and δ_k are the mean and covariance of the samples in class $i, (0 < i < C)$, and C is the number of individuals, respectively. $\bar{F} = (f_1, f_2, \dots, f_N)$ is the feature vector of a sample with N

features. The quadratic discriminant function (the special case of which is linear discriminant) is as follows [7],

$$D_i(\bar{F}) = (\bar{F} - \mu_i)^T . \delta_i{}^{-1} . (\bar{F} - \mu_i) + ln|\delta_i| - 2ln\pi_i, \quad (4.13)$$

where $(\bar{F} - \mu_i)^T . \delta_i{}^{-1} . (\bar{F} - \mu_i)$is Mahalanobis distance between $\bar{F}$ and μ_i. RDC is regularized form of discriminant classification by adding regularization values to class scatter matrix [7],

$$D^R{}_i(\bar{F}) = (\bar{F} - \mu_i)^T . \delta_i{}^{-1}(\rho, \gamma) . (\bar{F} - \mu_i) + ln|\delta_i(\rho, \gamma)| - 2ln\pi_i, \quad (4.14)$$

where ρ, γ are the regularization parameters ($\rho, \gamma \in [0,1]$) [7].

4.3.2 Fuzzy Nearest-Neighbour (FNN)

In the nearest neighbour classifier we assign a class to a probe sample which has the minimum distance to the gallery sample of the corresponding class and do not take the contribution of each labelled sample in attributing the class label which leads to misclassification when samples overlap [8]. However, in fuzzy nearest neighbour classification the fuzzy membership values per class are devoted to the sample and decision is made according to the maximum vote which increases the classification confidence. The class membership is calculated using the Euclidean distance between the sample and the class means prototype as follows [8],

$$FM_i(s) = \frac{\|s - M_i\|^{-\frac{2}{m-1}}}{\sum_{j=1}^{C} \|s - M_j\|^{-\frac{2}{m-1}}}, \quad (4.15)$$

where $FM_i(s)$ is the membership value of sample s in the i^{th} class, M_i is the mean of i^{th} class and C is the number of classes.

4.3.3 Sparse Classifier (SC)

Sparse classifier which is based on the sparse representation is general form of nearest neighbour technique and has a remarkable discriminative power [9]. The probe samples can be

represented sparsely as a linear combination of gallery samples of the same individual .The sparsest representation can be achieved by using the l_1-norm [9, 13]. The matrix of gallery samples is as follows,

$$G = [G_1, G_2, G_3, \dots, G_N], \tag{4.16}$$

$$G_i = \left[F_1^g, F_2^g, F_3^g, \dots, F_L^g\right], \tag{4.17}$$

where G_i is the matrix of gallery images of the i^{th} class and F_k^g is the feature vector of the k^{th} sample in G_i. N and L are the number of individuals and gallery samples per class, respectively. A probe sample is represented as follows [13],

$$F_i^p = G.C, \tag{4.18}$$

where $C = [0,0,\dots,0,c_1^i, c_2^i, \dots, c_L^i, 0,0,\dots,0]$ and c_j^i is the j^{th} coefficient associated with the i^{th}class.

In order to find the identity of the probe sample, we require to solve the l_1-norm minimization problem as follows [13],

$$(l_1): \quad \hat{C}_1 = argmin\|C\|_1 \qquad while\ F^p = G.C. \tag{4.19}$$

4.3.4 Extreme Learning Machine (ELM)

Feed forward neural networks are slow in learning and poor in scalability since they have to tune the weights iteratively via the gradient descent approach which is very slow. Extreme learning machine is a generalized form of single hidden layer neural networks with random hidden nodes where the weights of hidden layer are determined and do not require to be adjusted in iterations [10, 13]. Therefore, it is significantly faster than gradient descent which is an important issue in pattern recognition problems.

4.4 INTERPOLATION

In real word applications, the captured images usually have very low resolution which degrades the accuracy of the identification algorithm. In order to enhance the image resolution, one of the image interpolation techniques is required to enlarge the probe image. In this chapter we have employed the nearest neighbour, bilinear and bicubic interpolation methods to increase the image resolution before applying the face recognition algorithm on the image.

4.4.1 Nearest Neighbour Interpolation

This method is the fastest and simplest interpolation technique based on copying the grey value of the nearest neighbour of the reference pixel which leads to a larger image [11].

4.4.2 Bilinear Interpolation

In bilinear technique the intensity value of each unknown pixel is predicted using a weighted combination of the grey values of four neighbours. The resulting image is smoother than the nearest neighbour interpolated image [11].

4.4.3 Bicubic Interpolation

In order to find the intensity value of the interpolated pixels, we calculate the convolution of a cubic function with the 4×4 neighbourhood around the reference pixel [11].

4.5 EXPERIMENTAL RESULTS

4.5.1 Databases and Simulation Settings

In order to evaluate the performance of the proposed feature extraction and classification combinations, we utilize ORL and AR databases. Gabor filter has 8 scales, 5 orientations and $f_{max} = 0.25$. The value of m in FNN is equal to 2 and 0.1 is selected as the regularization factor in RDC, which are chosen experimentally through an exhaustive search.

1) *ORL:* AT&T or Olivetti research lab (ORL) database which consists of 400 images of 40 people (10 per individual) of size 112×92 with different time, scales and illumination

conditions, and various facial expressions (open/close eyes and smiling/not smiling) and up to 20 degree tilting and rotation [14]. We use the first 5 samples of each individual in the gallery and the rest 5 images in the probe dataset. Therefore, we have 200 images in the gallery set and 200 in the probe set.

2) *AR:* AR face database contains more than 3000 facial images of 136 people (76 men and 60 women) [15]. 26 images of each individual were taken in two sessions in two weeks (13 images per session). The database consists of eye aligned images of size 165×120 pixels with illumination variation, facial expression and partial occlusion [15]. We have employed non-occluded images 1 to 4 with different facial expressions in session 1 as the gallery and non-occluded images 2 to 4 in session 2, with appearance changes in time gap and facial expression and illumination variation, as the probe set.

4.5.2 Experiment I

In the first experiment we employed the proposed feature extraction and classification algorithms on the images of size 128×128 pixels in the mentioned datasets and the results are shown in Table 4.1 and 4.2 for ORL and AR databases, respectively.

Table 4.1. Identification accuracies (%) for 128×128 pixels images in ORL database with different combinations of feature extraction and classification techniques.

Classification / Feature Extraction	FNN	ELM	SC	RDC
Gabor-PCA	96	99.5	97.5	99
Gabor-LDA	98	99.5	98	99
Gabor-LSDA	96.5	99.5	97.5	96.5
Gabor-NPE	99	96.5	96.5	98
GLPQH- PCA	93	95	96.5	95
GLPQH- LDA	94.5	93.5	95	94.5
GLPQH- LSDA	92.5	90.5	91	93
GLPQH- NPE	88	91.5	92	89

Table 4.2. Recognition rates (%) for images of size 128×128 pixels in AR database with different combinations of feature extraction and classification techniques.

Classification / Feature Extraction	FNN	ELM	SC	RDC
Gabor-PCA	83	96.33	92.33	93.67
Gabor-LDA	96.33	96.33	95.67	96.67
Gabor-LSDA	74.67	96.33	86	75.67
Gabor-NPE	78	96.33	89	79
GLPQH- PCA	89.33	87.67	93	92
GLPQH- LDA	93	90	92	92.33
GLPQH- LSDA	48.67	81.33	55.67	47.33
GLPQH- NPE	45.33	87	55	44.33

We repeat the experiment for low resolution images of size 16×16 pixels and Table 4.3 and 4.4 illustrate the results for ORL and AR databases, respectively.

The results show that utilizing PCA and LDA as dimensionality reduction techniques leads to better results for both databases. ELM has the best average performance for both databases. FNN shows the least stability versus resolution reduction for all feature extraction methods.

Table 4.3. Identification Accuracies (%) for images of size 16×16 pixels in ORL database with different combinations of feature extraction and classification techniques.

Classification / Feature Extraction	FNN	ELM	SC	RDC
Gabor-PCA	94	95.5	97	97
Gabor-LDA	95.5	95.5	93.5	96
Gabor-LSDA	94	95.5	94.5	95.5
Gabor-NPE	90.5	94	94	94
GLPQH- PCA	96.5	97	95.5	96
GLPQH- LDA	96.5	97	95.5	96
GLPQH- LSDA	96.5	96	95	96.5
GLPQH- NPE	96.5	97.5	96.5	93.5

Table 4.4. Recognition rates (%) for 16×16 pixels images in AR database with different combinations of feature extraction and classification techniques.

Classification / Feature Extraction	FNN	ELM	SC	RDC
Gabor-PCA	64.33	79.33	72.33	72.67
Gabor-LDA	68	67.33	68.67	70.33
Gabor-LSDA	49.67	67	51	50.33
Gabor-NPE	52.67	67.68	51.67	52.33
GLPQH- PCA	59.33	67.67	65.33	69.33
GLPQH- LDA	71.67	69.33	70.33	67.33
GLPQH- LSDA	54.33	69	61	54.33
GLPQH- NPE	57.33	68.33	61.33	56.67

4.5.3 Experiment II

In this experiment we apply three interpolation techniques on 16×16 images to enhance the resolution to the size of 128×128 and then apply the face recognition techniques. Figures 4.3 and 4.4 illustrate the results on ORL and AR databases, respectively.

Nearest neighbour interpolation performs better than the other two methods for the ORL database and with Gabor technique the results are very close to the recognition rates in Table 4.1.

Figure 4.4 shows that the bicubic and bilinear methods performances are comparable, but nearest neighbour interpolation outperforms those two approaches and ELM shows almost consistent performance in combination with all feature extraction techniques.

4.6 CONCLUSION

In this chapter we have evaluated the performance of face recognition algorithms on low resolution images of the ORL and AR databases and the effect of interpolation techniques to enhance their results. We have used different combinations of feature extraction and classification approaches. Gabor and its combination with LPQ histogram (GLPQH) are utilized as image descriptors and to deal with the high dimensionality problem we have used PCA, LDA, LSDA and NPE methods. Extreme learning machine, sparse classifier, fuzzy nearest neighbour and

regularized discriminant classifiers are employed to assign the class labels. The results of experiments show that ELM and RDC have acceptable performance and stability against resolution reduction, especially in combination with Gabor technique with LDA or PCA dimensionality reduction. Nearest neighbour interpolation outperforms bilinear and bicubic interpolations in order to enhance the low resolution image.

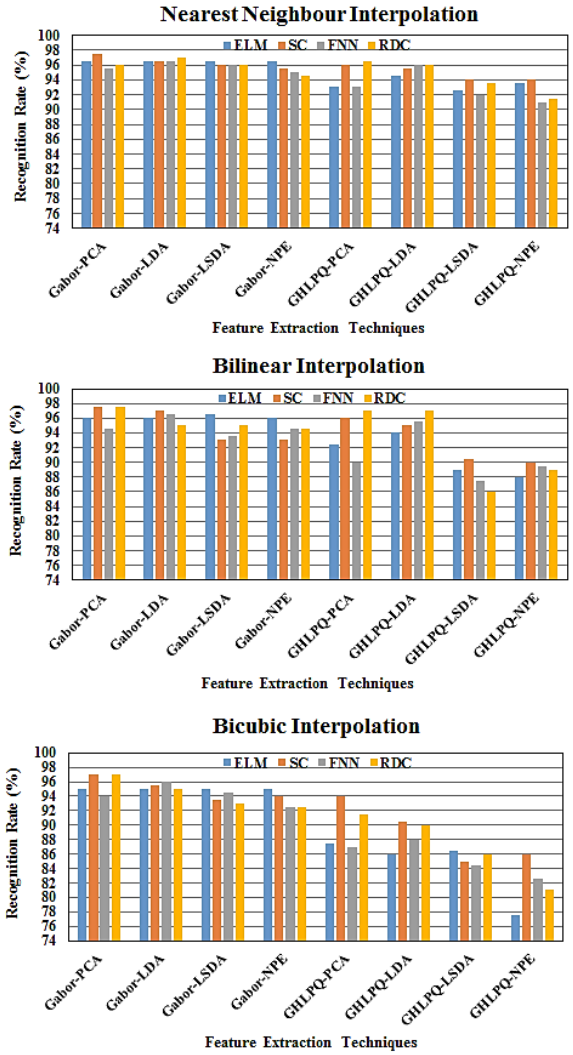

Figure 4.3. Recognition accuracies (%) for different feature extraction and classification combinations on the interpolated images in ORL database (from 16 × 16 pixels to 128 × 128 pixels) with three interpolation techniques; nearest neighbour, bilinear and bicubic.

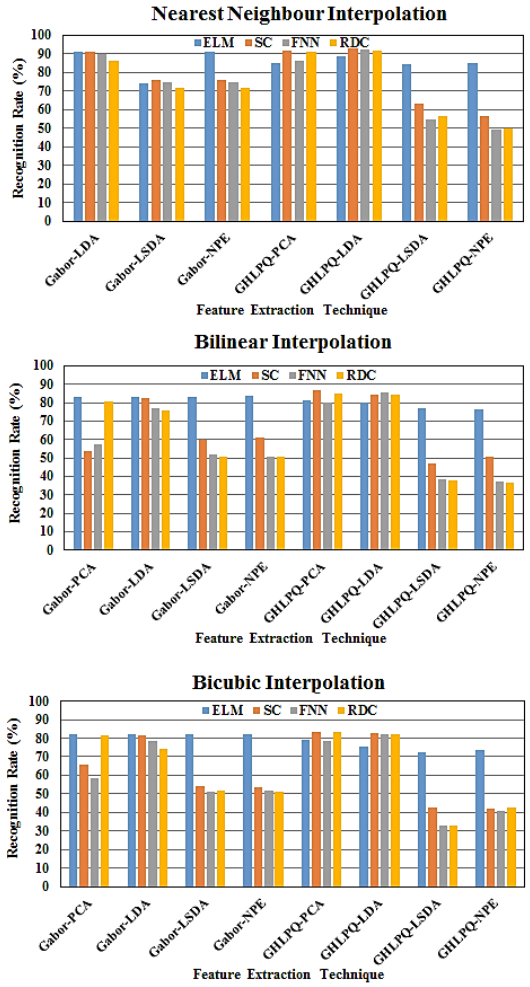

Figure 4.4. Recognition accuracies (%) for different feature extraction and classification combinations on the interpolated images in AR database (from 16×16 pixels to 128×128 pixels) with three interpolation techniques; nearest neighbour, bilinear and bicubic.

4.7 REFERENCES

[1] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression," *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol. 36, pp. 1169-1179, Jul. 1988.

[2] M. Turk, A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*., vol. 3, pp. 71-86, Winter 1991.

[3] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of Optical Society of America A.*, vol. 14, pp. 1724-1733, 1997.

[4] D. Cai, X. He, K. Zhou, J. Han and H. Bao "Locality sensitive discriminant analysis," in *Proc. IJCAI'07*, 2007, pp. 708-713.

[5] X. He, D. Cai, S. Yan, H. J. Zhang, "Neighborhood Preserving Embedding," in *Proc. ICCV'05*, 2005, p. 1208-1213.

[6] C.H. Chan, M. A. Tahir, J. Kittler, M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 35, pp. 1164-1177, May. 2013.

[7] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, pp. 165-175, Jul. 1989.

[8] J. M. Keller, M. R. Gray, J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 15, pp. 580-585, Jul-Aug. 1985.

[9] J. Wright, A.Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210-227, Feb. 2009.

[10] G. B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 45, pp. 513-529, Oct. 2012.

[11] S. M. Metev, V. P. Veiko, *Practical Image and Video Processing Using MATLAB*, 1st ed., Hoboken, New Jersey: John Wiley & Sons, 2011.

[12] S. Nikan, M. Ahmadi, "Study of the effectiveness of various feature extractors for human face recognition for low resolution images," in *International Conference on Artificial Intelligence and Software Engineering (AISE'14)*, Phuket, Jan 2014, pp. 1-6.

[13] S. Nikan, M. Ahmadi, "Effectiveness of various classification techniques on human face recognition," in *Proc. High Perf. Computing and Simulation (HPCS)*, Jul 2014, pp. 651-655.

[14] The AT&T Laboratories Cambridge website, Available (online) at: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

[15] A. Martinez, R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998. Available (online) at: < http:// www2.ece.ohio-state.edu/~aliex/ARdatabase.html >.

Chapter 5

Recognition of Human Faces under Different Degradation Conditions

Face recognition is one of the most remarkable image analysis and biometric identification techniques which is utilized in a wide range of applications such as human-computer interaction, law enforcement, video surveillance, and forensic investigation. An automatic face recognition system has a significant capability in distinguishing individuals from their face images without the human cooperation requirement. It is composed of three main stages, image preprocessing which is employed in order to enhance the appearance of face image and reduce the irrelevant effects, which might lead to misclassification of the image, feature extraction which is carried out by utilizing an image descriptor to derive distinctive characteristics of image and classification which attributes the most probable identity label, among a large number of subject classes, to an unknown individual [1]. Although there exist many commercial face recognition algorithms which are highly accurate in constraint environment, it is still an extremely challenging computer vision research area due to the sensitivity of identification accuracy to the appearance variations which are occurred in uncontrolled conditions, such as illumination and head pose variation, facial expression, partial occlusion, blur and low resolution problem and aging effect. In this chapter, illumination variation and partial occlusion will be addressed in details which are very crucial problems in real world applications and mislead face recognition algorithm considerably.

5.1 ILLUMINATION VARIATION CHALLENGE

Illumination variation has a remarkable effect on the face appearance and the intra-subject changes in the images belonging to one person at various lighting conditions might be more significant than the inter-personal differences between two different individuals [2]. Therefore, it can reduce the accuracy of face recognition system substantially. Within the past few decades different strategies have been applied in order to cope with the image degradation caused by lighting condition, which are divided into three categories; (i) illumination insensitive image processing, which utilizes gray level transformations or gradient image to reduce the illumination effect, or extracts the reflection component and removes the lighting effect, (ii) illumination invariant image descriptors to extract image features which are insensitive to lighting effect and (iii) block-based pattern recognition techniques which reduce the influence of degraded image sub regions on the identification process [1, 3], where combination of all categories improves the recognition performance.

5.1.1 Illumination Insensitive Image Processing

There exist a variety of image processing approaches to normalize the face image and remove the lighting influence. These techniques are divided into three categories of approaches as follows.

5.1.1.1 Intensity level transformation

In this category of illumination insensitive image processing, a linear or non-linear transformation is applied on the gray level images to map the intensity and correct the illumination effect to some extent [1]. Gamma correction (GC) [4] and histogram equalization (HE) [5] are two examples of non-linear gray-level transformations which correct the uneven intensity distribution of lighting effect in the image [1]. In GC each image pixel is replaced with its gray value to the power of γ as follows [4],

$$I = \begin{cases} I^{\gamma} & for \quad \gamma > 0 \\ \log(I) & for \quad \gamma = 0 \end{cases}, \quad where \;\; \gamma \in [0\ 1]. \tag{5.1}$$

HE is a nonlinear transformation of the image intensity values, k, which enhances the image

contrast by spreading the intensity distribution of image in order to create a new image with flat histogram [5] as follows,

$$T(k) = floor(L-1)\sum_{n=0}^{k} h(n), \tag{5.2}$$

where $h(n)$ is image histogram and L is the maximum intensity level.

In order to further reduce the effect of lighting, the modified version of HE in the form of an orientated local histogram equalization (OLHE) was proposed in [6], where, the illumination insensitive face representation is captured while the rich face information on the edge orientations are preserved.

Difference of Gaussian (DoG) [7] is another gray scale transformation method which is a band-pass filter, consists of two low-pass filters with standard deviations $\sigma_1$andσ_2, to enhance the edges in image. DoG at pixel position (x, y) is as follows,

$$DoG(x,y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}}. \tag{5.3}$$

5.1.1.2 Gradient based techniques

Gradient-based preprocessing techniques find the illumination invariant representation of image using the fact that the illumination component of image,$L(x, y)$, changes slowly, in comparison with the abrupt variations in the reflection component, $R(x, y)$ [1]. In [8], based on the reflection image model, illumination insensitive representation of face image, I, is derived using arctangent of the ratio between directional derivatives of face image, which is called gradient face (GF),

$$I(x,y) = R(x,y) \times L(x,y), \tag{5.4}$$

$$GF = tan^{-1}\left[\frac{\frac{\partial I(x,y)}{\partial y}}{\frac{\partial I(x,y)}{\partial x}}\right] = tan^{-1}\left[\frac{L(x,y)\times\frac{\partial R(x,y)}{\partial y}}{L(x,y)\times\frac{\partial R(x,y)}{\partial x}}\right] = tan^{-1}\left[\frac{\frac{\partial R(x,y)}{\partial y}}{\frac{\partial R(x,y)}{\partial x}}\right]. \tag{5.5}$$

Xu et al. improved their technique [9] by extracting ratio of wavelet filtering of y- derivative of image over x-derivative to obtain the multi-resolution illumination invariant analysis of image.

The psychological law which was proposed by Ernst Weber in 1834 implies that the ratio between a noticeable variation in a stimulus and its original value is a constant [10]. Based on the Weber's law, Wang et al. calculated the illumination insensitive image representation by using the tangent inverse of the ratio between the local intensity variation, which is computed using the Laplace operator, and is called "Weber-face" (WF) [11],

$$WF = tan^{-1}\left[\alpha \sum_{i\in A}\sum_{j\in A}\frac{I(x,y)-I(x-i\Delta x,y-j\Delta y)}{I(x,y)}\right] = tan^{-1}\left[\alpha \sum_{i\in A}\sum_{j\in A}\frac{L(x,y).[R(x,y)-R(x-i\Delta x,y-j\Delta y)]}{L(x,y).R(x,y)}\right]. \quad (5.6)$$

Subspace learning based approaches are also very popular techniques in illumination insensitive image preprocessing. Tzimiropoulos et al. proposed principal component analysis (PCA) [12], linear discriminant analysis (LDA) [13], locally linear embedding (LLE) [14] and laplacian eigenmaps (LE) [15] learning methods to obtain a robust image representation against illumination variation based on the gradient orientations of image instead of the pixel gray values. The cosine kernel is employed to calculate the image correlation which leads to robustness versus noise and outliers [16].

5.1.1.3 Reflection component estimation

Another category of illumination invariant preprocessing techniques is based on the estimation of reflectance component of the image [1]. Self-Quotient image (SQI), proposed by Wang et al. [17], is the ratio between the input image and its smoothed version, using a smoothing kernel S, based on the concept of Quotient method,

$$SQI = \frac{I}{\hat{I}} = \frac{I}{S*I}. \quad (5.7)$$

Wavelet transform has been used as a viable technique for illumination invariant image processing. Cao et al. [18] applied a de-noising method on the high frequency coefficients of the wavelet transform and the reflectance component has been considered as noise. Thresholding was proposed to remove noise, thus the luminance component was obtained and subtracted from the

logarithmic image to extract illumination invariant component.

The luminance and reflection components were separated using the reflection model in logarithmic form in the proposed algorithm in [19],

$$\log\big(I(x,y)\big) = \log\big(R(x,y)\big) + \log\big(L(x,y)\big). \tag{5.8}$$

However, Baradarani et al. utilized a normalized form of logarithm function to enhance the illumination, which can be tuned by two parameters, ξ and ϵ, as follows,

$$\hat{I}_{pro} = \xi \frac{\log\big(I(x,y)\big)}{\sqrt{\log\big(I(x,y)\big)-\epsilon}}. \tag{5.9}$$

In order to remove the illumination effect, they employed a double-density dual-tree complex wavelet transform (DD-DTCWT) with 32 directional wavelets to decompose the frequency sub-bands of the illumination enhanced image and the high frequency components are thresholded by the minimum frequency coefficient in every row of the high frequency sub bands. If the value of the frequency coefficients are greater than the threshold value they are set to zero. The luminance component is preserved by thresholding the high frequency sub bands and reconstructing by inverse DD-DTCWT. The reconstructed image is subtracted from the logarithmic image to extract the reflection component. In order to further reduce the effect of illumination, low-pass filtering and edge amplifying were also applied on the reconstructed image. DD-DTCWT is utilized in this stage to extract the features of the illumination insensitive image and extreme learning machine is applied as the classification technique [19]. Because of the poor scalability and slow learning of feed forward neural networks, extreme learning machine [20] is used. Therefore, the iterative weight adjustment is not required and it is extremely faster than gradient descent technique [20].

Discrete cosine transform (DCT) is another image transformation to separate the frequency sub bands. Chen et al. applied DCT on the image logarithm [21] and by eliminating the low frequency DCT coefficients, which are related to lighting effect, the illumination variation was compensated. Chen et al. applied the logarithmic form of reflection model as follows,

$$\log\big(I(x,y)\big) = \log\big(R(x,y)\big) + \log\big(L(x,y)\big) = \log\big(R(x,y)\big) + \log\big(\hat{L}\big) + \epsilon(x,y), \quad (5.10)$$

where $L(x,y)$ is the incident illumination component and $\hat{L}$ is the desired uniform illumination, which affects the image pixels uniformly, and $\epsilon(x,y)$ is the compensation term, which is the difference between normalized and original illumination effects. Due to the fact that luminance component varies slowly compared to the reflectance, it is the low frequency component of image [21]. Chen et al. employed images containing only the face area since hair is a low frequency image component and varies by time changing, thus is not a stable feature. DCT is a redundancy reduction based on a set of orthogonal cosine functions [21],

$$D(p,q) = \alpha(p)\alpha(q)\sum_{m=0}^{R-1}\sum_{n=0}^{C-1} I(m,n)\cos\left[\frac{(2m+1)p\pi}{2R}\right]\cos\left[\frac{(2n+1)q\pi}{2C}\right], \quad (5.11)$$

$$where \begin{cases} \alpha(p) = 1/\sqrt{R} & if\ p = 0. \\ \alpha(p) = \sqrt{2/R} & if\ p = 1,2,\dots,R. \\ \alpha(q) = 1/\sqrt{C} & if\ q = 0. \\ \alpha(q) = \sqrt{2/C} & if\ q = 1,2,\dots,C. \end{cases} \quad (5.12)$$

In order to discard the illumination effect the authors of [21] applied DCT on the logarithmic image and set the low frequency coefficients to zero. DCT of image results in a matrix of frequency coefficients of the same size as image. They illustrated that the coefficients with large values of standard deviation, corresponding to the low frequency components, are located in the top left corner of DCT coefficients matrix. Thus, an appropriate number of coefficients, in the manner which is shown in Fig 5.1, are discarded to reduce the illumination effect on the image [21].

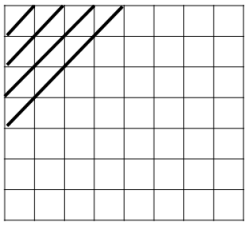

Figure 5.1. The manner in which the DCT coefficients are discarded.

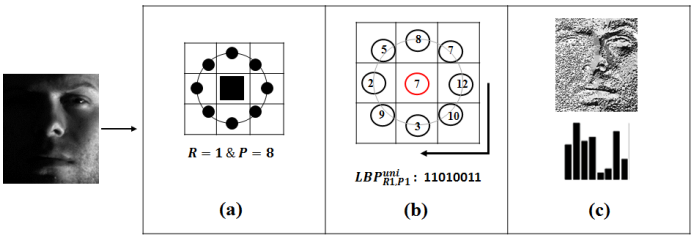

Figure 5.2. LBP: (a) selecting neighbourhood size, (b) comparing each pixel and its neighbors and (c) utilizing pixels decimal values to make image histogram.

5.1.2 Illumination Invariant Image Descriptor

Local binary pattern (LBP) is a histogram based illumination insensitive image descriptor [22]. Based on the comparison between the intensity values of the neighbour pixels, this technique is robust against monotonic image variations with fast and easy execution and no learning requirement [22, 23]. As shown in Fig 5.2 LBP extracts texture features of image by comparing each pixel with its neighbours, where R is the radius of neighborhood and P is number of neighbours to compare with the centre pixel. Therefore, a P-bit binary pattern for each image pixel is constructed, as illustrated in Fig 5.2, and thus we have 2^P decimal values for image pixels. The corresponding decimal values are used to make the image histogram with 2^P bins.

However, utilizing the combination of illumination insensitive image processing and illumination invariant feature extraction techniques could lead to better performance. In the approach proposed by Tan and Triggs [24], a series of preprocessing techniques consists of GC [4], DoG filtering [7], masking and equalization of image contrast using HE [5], were utilized along with a kernelized illumination insensitive local descriptor which is the modification of LBP, called local ternary pattern (LTP).

5.1.3 Block-Based Illumination Invariant Pattern Recognition

Holistic-based approaches, compared with block-based image analysis have less stability against intra personal variations. By employing the whole face image, appearance changes which occur in some regions of the image might influence all of the extracted features. However, lighting effect may influence certain areas of human face. In such cases, local based strategies are more adequate to cope with the local appearance changes where local face characteristics are extracted

separately. Thus, degraded regions of face affect only the features of the corresponding sub region. In addition, despite the stability of local based approaches against intensity asymmetry caused by shadow in side-lighting, holistic based methods do not tackle that problem properly [1]. Nabatchian et al. proposed a Max filtering technique on the logarithmic reflection model of image [25] in order to extract the luminance component. The output of maximum filter is subtracted from the logarithm image, which results in the illumination insensitive image representation as shown in Fig 5.3. They applied a local based technique by dividing the illumination insensitive image into non-overlapping image partitions and principal component analysis (PCA) was utilized to extract local features from different sub blocks. Support vector machine (SVM) [26] was applied as the local classifier to assign class labels to the image partitions and they employed majority voting scheme to combine local decisions to find the best match in gallery set for the corresponding probe image. Subsequently, the same feature extraction and classification techniques were applied on the whole face area as a holistic based face recognition and its result was taken into account in the voting scheme. In order to vary the contribution of different sub blocks in the classification of the probe image, the authors of [25] assigned an adaptive weight to each sub block. The adaptive weight is the fusion of three different weights. First, the difference between the mean intensity value of pixels in each sub block of the probe image and the corresponding sub block in the canonical image is calculated, which shows the influence of illumination on that sub region. The canonical image is the mean image of all gallery samples which are having a good level of illumination. Smaller difference shows better illumination condition of the probe image sub block which means its local result should have more effect on the majority voting,

$$d_{sb}^{P} = m_{sb}^{P} - m_{sb}^{canonical}, \tag{5.13}$$

where m_{sb}^{P} and $m_{sb}^{canonical}$ are mean gray values of the sub blocks in the probe and canonical

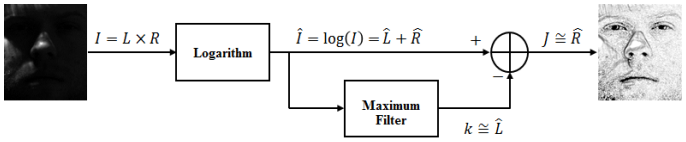

Figure 5.3. Illumination insensitive image representation using Max filter [25].

images, respectively [25]. The second proposed weight is based on the entropy or the information property of the probe image sub block which is calculated as follows,

$$En_{sb}^{P} = -\sum_{i=1}^{N_t} p_i \log(p_i), \tag{5.14}$$

where N_t shows the total number of gray levels in the probe image sub block and p_i is probability of the i^{th} intensity level [25]. The third proposed weight in [25] is calculated using the mutual information between the sub block of the probe image and the corresponding sub region in the reference image. The reference image is the mean image of the available gallery samples for the class of subject which is assigned to the sub block of the probe image. The larger value of the mutual information indicates the greater similarity between two images. The mutual information of two random variables X and Y with marginal probability distribution functions, $p_1(x)$ and $p_2(y)$, and joint probability distribution function $p(x, y)$, is calculated as follows [25],

$$\mu I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log\left(\frac{p(x,y)}{p_1(x).p_2(y)}\right). \tag{5.15}$$

The adaptive weight which is assigned to each sub block in the probe image is calculated by combining the aforementioned local weights, which are normalized between 0 and 1 [25],

$$AWG_{sb}^{P} = \mu I_{sb}^{P} \times (En_{sb}^{P})^2 \times \left(d^{max} - d_{sb}^{P}\right). \tag{5.16}$$

Fig 5.4 illustrates some images of the Yale B (YB) database [27], which contains 5,760 images for 10 individuals with 576 viewing conditions (9 poses and 64 illumination conditions per pose). Their illumination insensitive representations using the aforementioned techniques, DCT [21], DoG [7], gradient faces [8], LBP [22], maximum filter [25], SQI [17], Weber faces [11] and wavelet denoising [18], are shown in Fig 5.4.

Moreover, the database is divided into five subsets corresponding to the angle between the light source and the camera axis, α; subset 1 consists of 7 images per subject, ($\alpha < 12^{\circ}$) and subsets 2

to 5 contain 12, 12, 14 and 19 images per subject, with ($20^{\circ} < \alpha < 25^{\circ}$), ($35^{\circ} < \alpha < 50^{\circ}$), ($60^{\circ} < \alpha < 77^{\circ}$) and ($\alpha > 78^{\circ}$), respectively. Table 5.1 shows the recognition accuracies of the above methods when subset 1 is used as the gallery set and subsets 2-5, are employed as four probe sets [25].

Table 5.1. Accuracy of the face identification algorithms for subsets 2-5 of the YB database (%).

Method	Subset2	Subset3	Subset4	Subset5
LBP[22]	100	97.6	65.2	44.4
DoG[7]	100	97.4	98.7	93.4
Weberface[11]	100	100	96.4	96.8
SQI[17]	100	100	96.4	97.9
DCT[21]	100	100	99.82	98.29
gradientfaces[8]	100	100	99.28	100
Max+PCA/SVM+WVS[25]	100	100	99.29	100
wavelet denoising[18]	100	100	100	100

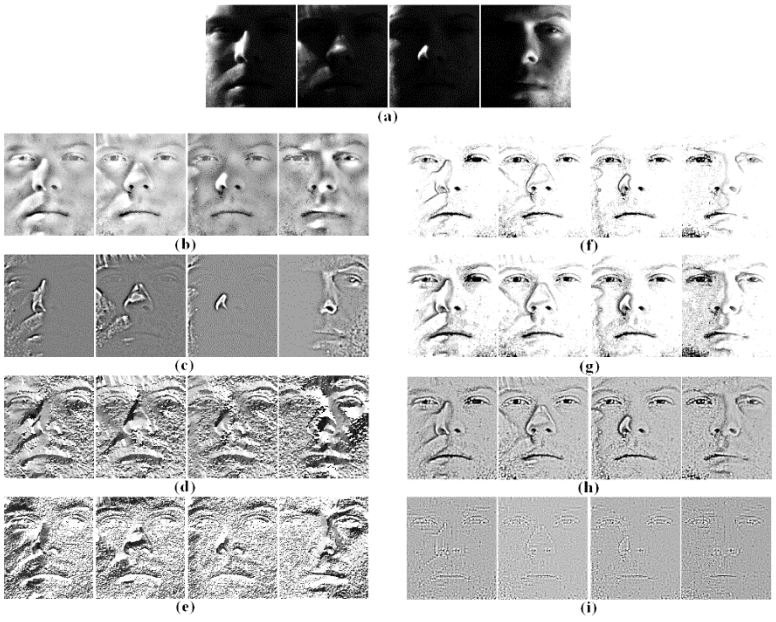

Figure 5.4. Samples of YB database: (a) original images, and illumination insensitive images using: (b) DCT [21], (c) DoG [7], (d) gradient faces [8], (e) LBP [22], (f) Max filter [25], (g) SQI [17], (h) Weber faces [11] and (i) wavelet denoising [18].

5.2 PARTIAL OCCLUSION CHALLENGES

Occlusion is another crucial challenge in face recognition which occurs in real-world applications such as video surveillance in security controlling systems where the criminals try to cover some parts of their faces or occlusion caused by other people or objects in a crowded scene. Thus, the face area is occluded by an obstacle such as sun glasses, scarf, hat or another face or object [28]. Therefore, some parts of facial information are lost which can affect the accuracy of face identification and depending on the regions which are occluded and the size and texture of occlusion, accuracy reduction is different. In some cases the non-occluded portion of image does not contain sufficient distinctive information and the face identification is difficult or impossible [29]. In holistic based face description techniques such as PCA [12] and LDA [13], the whole face is applied to extract features and the occluded region of the face can affect the whole feature vector [30]. While, partial occlusion corrupts some areas of face and local based descriptors are good alternatives in order to tackle the partial occlusion problem where the face image is divided into small patches and local degradation affects the local features of the corresponding occluded regions. The effect of occluded areas on the recognition system can be reduced by detecting the occlusion [30]. Occluded image partitions can be determined by training the face recognition system using both occluded and non-occluded samples. Gabor filter [31], with PCA as a dimensionality reduction technique, was applied in [30] to extract the characteristics of image sub regions and train the face recognition system by occluded and non-occluded samples. SVM [26] was utilized to attribute the occluded or non-occluded labels to sub regions of the probe image and LBP [22] extracted features of non-occluded partitions. However, in real-world and in uncontrolled situations, the available database does not contain occluded image samples and in some cases we have only one training sample per individual available. In such cases we need to apply an information theory technique in order to detect the occluded parts according to the information content of the image partitions. As soon as we categorize the image patches as occluded or non-occluded, there are two categories of face recognition approaches available to deal with partially occluded images. In the first group, we can reconstruct the corrupted parts by a reconstruction technique and apply a face recognition technique on the repaired image. Second category consists of methods where the occluded parts are excluded from the identification process and face recognition is carried out using the information extracted from the non-occluded

areas, or local weights are assigned to image sub regions and the influence of corrupted regions, on the face recognition process, is reduced by devoting lower weights to those regions [32]. In this chapter, the second category of identification approaches is considered.

5.2.1 Excluding Occluded Face Regions or Reducing Their Effect

In this category of occluded face recognition techniques, the image completion or reconstruction is not carried out due to the fact that it is not fast enough and computationally cost effective and requires a large set of training samples. Therefore, the recognition is based on a local based technique and excluding the occluded pixels or minimizing their effect on the recognition algorithm [32]. A local probabilistic subspace method was used by Martinez [33] for occluded face recognition with one sample per subject by dividing image into sub blocks and finding the subspace for each sub region in order to project the sub blocks in the training set to the eigenspaces. For a probe image, the k^{th} sub block, P_{sb}^{k}, was projected to the corresponding eigenspace and the local probability for each sub region was calculated using the projected sub block of probe image, $\hat{P}_{sb}^{k}$, and mean, $\mu_{j,k}$, and covariance, $\Psi_{j,k}$, of the subspace as follows,

$$LocProb_{j,k} = \left(\hat{P}_{sb}^{k} - \mu_{j,k}\right).\Psi_{j,k}.\left(\hat{P}_{sb}^{k} - \mu_{j,k}\right), \tag{5.17}$$

where $k = 1,2,\dots,N_{sb}$ and $j = 1,2,\dots,N_C$. N_{sb} and N_C are the number of sub blocks and number of classes, respectively. The local probabilities were taken into account and summed up and the maximum of its argument calculated as the class label of probe image [33],

$$Class = argmax_j\left\{\sum_{k=1}^{N_{sb}} LocProb_{j,k}\right\}. \tag{5.18}$$

J. Kim and et al. proposed a local based face recognition algorithm [34] by dividing image into a finite number of non-overlapping sub blocks and the occluded sub regions were detected using a nearest neighbour classifier on the PCA sub-space and then identification based on a selective local non-negative matrix factorization (LNMF) was carried out. LNMF bases in the occluded regions are detected by calculating the portion of total energy which belongs to an occluded region for each LNMF, as follows,

$$E_i = \frac{\sum_{x,y \in W} B_i^2(x,y)}{\sum_{x=1}^{R} \sum_{y=1}^{C} B_i^2(x,y)}, \quad i = 1,2, \dots, N_b, \tag{5.19}$$

where $B_i(r,c)$ is the value of the i^{th} LNMF at (x,y) image coordinate, W is the detected occluded face region, $R \times C$ is the size of image and N_b is the total number of LNMF bases. If the energy value is greater than a threshold, that basis is considered as an occluded LNMF. Data projection is carried out using only the non-occlude bases, the energy value of which is smaller than the threshold. For a probe image the set of corresponding occlusion-free bases is $B = [b_1, b_2, \dots, b_M]$, where its pseudo inverse, B^{PI}, is used to project the probe image and gallery set to the occlusion-free LNMF subspace [34],

$$\hat{P} = B^{PI} \times P, \tag{5.20}$$

$$\hat{G_i} = B^{PI} \times G_i, \tag{5.21}$$

where P and G_i are probe and i^{th} gallery images, respectively, and $\hat{P}$ and $\hat{G_i}$ are their projection coefficients in the LNMF subspace [34]. Coefficients of the probe and gallery images are compared to find the closest match.

Kanan et al. in [35] proposed pseudo Zernike moment (PZM), to extract features of image sub regions. As shown in Fig 5.5 the sample face image of size $M \times M$ pixels from the AR database, which contains 3536 facial images of 136 people (76 men and 60 women) affected by partial occlusion (scarf or sun glasses) [36], is partitioned into equal-sized and non-overlapping sub blocks of $w \times w$ resolution. Each sub region is bounded inside a unit circle and the sub block coordinates, x_i and y_j , are mapped to the polar form,

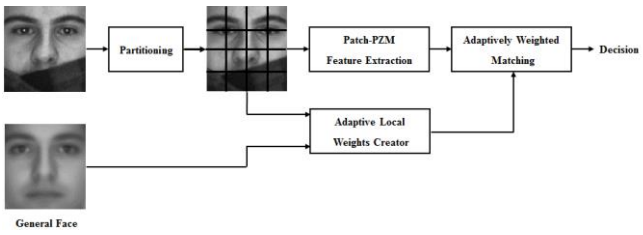

Figure 5.5. Block diagram of the proposed algorithm in [35].

$$\begin{cases} x_i = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{w-1} i, & i = 0, 2, \dots, w-1. \\ y_j = +\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{w-1} j, & j = 0, 2, \dots, w-1. \end{cases} \tag{5.22}$$

The PZMs, based on a set of orthogonal pseudo Zernike polynomials are extracted from image sub block, $I_{r,c}(x_i, y_j)$, at location (r, c) as follows [35],

$$PZM_{p,q}^{r,c}\left(I(x_i, y_j)\right) = \frac{p+1}{\pi\lambda(w)} \sum_{j=0}^{w-1} \sum_{i=0}^{w-1} Z_{p,q}^{*}(x_i, y_j) \times I(w(r-1) + x_i, w(c-1) + y_j), \tag{5.23}$$

where $r, c = 1, 2, \dots, M/w$, $\lambda(w) = \frac{w^2}{2}$ is the normalization factor employed for mapping to the unit circle and $Z_{p,q}^{*}(x, y)$ is the complex conjugate of the pseudo Zernike polynomial. Finally, the feature vector of the face image is constructed by concatenating the local PZMs of image sub blocks. The distance between the probe and gallery images, P and G, respectively, is as follows [35],

$$D(P(r,c), G(r,c)) = \sum_{r=1}^{M/w} \sum_{c=1}^{M/w} AWM_{r,c}(a,b) \times \left\{ \sum_{p=0}^{p_{max}} \sum_{q=0}^{p} \left[\left| PZM_{p,q}^{r,c}\left(P(x_i, y_j)\right) \right| - \left| PZM_{p,q}^{r,c}\left(G(x_i, y_j)\right) \right| \right]^2 \right\}, \tag{5.24}$$

$$AWM_{r,c}(a,b) = (AOM_{r,c}(a,b))^2 . CM_{r,c}(a,b), \quad a, b = 1, 2, \dots, W, \tag{5.25}$$

where $AOM_{r,c}(a,b)$ is the adaptive occlusion map which shows the difference between PZM features of the sub region in the probe image and the corresponding sub region in the general image. Where, the general image is the average of all gallery samples. Also, $CM_{r,c}(a,b)$ is the contribution map of the probe image sub block according to the information content of that sub region which is obtained by calculating the local entropy value in the image sub block using (5.14). The values of CM and AOM are normalized and between 0 and 1 [35].

A sparse representation was proposed by Wright et al. [37], where the choice of feature transformation is not critical anymore but it requires many sample images per class. A general classification was proposed based on sparse representation. The probe sample can be represented sparsely as a linear combination of gallery samples of the same class [37]. The matrix of gallery samples is as follows,

$$\bar{G} = [\bar{G}_1, \bar{G}_2, \bar{G}_3, \dots, \bar{G}_N], \tag{5.26}$$

$$\bar{G}_i = \left[G_1^i, G_2^i, G_3^i, \dots, G_{N_{G_i}}^i\right], \tag{5.27}$$

where $\bar{G}_i$ is the matrix of gallery samples of the i^{th} class and G_k^i is the feature vector of the k^{th} sample in $\bar{G}_i$. N_C and $N_{\bar{G}_i}$ are the number of classes and gallery samples per i^{th} class, respectively. A probe image can be represented as follows [37],

$$\bar{P} = \bar{G} \times \bar{C}, \tag{5.28}$$

$\bar{C} = [0,0,\dots,0,c_1^i, c_2^i, \dots, c_L^i, 0,0,\dots,0]$, where c_j^i is the j^{th}coefficient associated with the i^{th} class and L is the number of gallery samples per i^{th} class. In order to find the coefficient vector of the corresponding probe image, we need to solve the l_1-minimization problem as follows,

$$(l_1): \quad \hat{C}_1 = argmin\|\bar{C}\|_1 \qquad while\ \bar{P} = G \times C. \tag{5.29}$$

Wright et al. showed the important factor to recover the sparse representation and correctly assign the class label to the probe image is the dimension of the feature space which should be large enough and the l_1-norm solution being sufficiently sparse, thus the selection of feature space is not considerable. The robustness versus occlusion was undertaken by the assumption that the error occurred by partial occlusion has a sparse basis due to the fact that it affects just a fraction of image pixels. Therefore, the linear model in (5.28) is extended as follows [37],

$$\bar{P} = (\bar{G} \times \bar{C}) + e, \tag{5.30}$$

where e is the error vector and a small portion of its elements are nonzero. The location and magnitude of error which is caused by occlusion is unknown. The equation (5.30) is modified and the sparsest solution is found by solving the l_1-minimization as follows [37],

$$\bar{P} = [\bar{G}, I]\begin{bmatrix}\bar{C}\\ e\end{bmatrix} = \hat{G} \times V, \tag{5.31}$$

$$(l_1): \quad \hat{V}_1 = argmin\|V\|_1, \qquad while\ \ \bar{P} = \hat{G} \times V, \tag{5.32}$$

where $\hat{V}_1 = [\hat{C}_1, \hat{e}_1]$ and the occlusion is removed by setting $\bar{P}_r = \bar{P} - \hat{e}_1$.

Nikan and Ahmadi proposed a local based face recognition under occlusion by using one gallery sample per individual in the database [23]. Fig 5.6 illustrates the block diagram of their proposed algorithm. The face image is partitioned into equal-sized and non-overlapping sub blocks. The accuracy of the face recognition algorithm can be affected by the size of image sub blocks. Utilizing very large sub blocks is not computationally cost effective and does not satisfy the block-based idea. However, very small image partitions cannot provide satisfactory discriminative information [38].

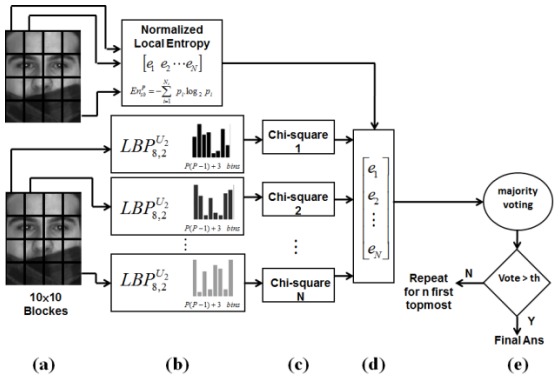

Figure 5.6. Proposed algorithm in [23]: (a) Image sub division, (b) LBP on each sub block, (c) local nearest neighbour classifiers, (d) local entropy calculation and (e) majority voting scheme and thresholding maximum vote to find final decision.

LBP descriptor which is robust to rotation, scaling and illumination variation, is applied on the image sub blocks to extract local texture features. LBP procedure is carried out on each image partition and binary patterns are employed to construct local histograms for image sub blocks. In order to reduce the feature dimensionality, Nikan and Ahmadi took only uniform binary patterns into account [23]. The binary code is uniform if there are at most two bitwise transitions from 0 to 1 or 1 to 0. Each local histogram contains $P(P-1)+2$ bins for uniform and 1 bin for all non-uniform patterns [22]. Local histograms of probe image sub regions are compared with the histograms of corresponding sub blocks in the gallery samples and classified using local Chi-square based nearest neighbour classifier which is one of the most accurate histogram similarity metrics as follows [7, 22],

$$\mathcal{X}^2\left(P^{sb}, G^{sb}\right) = \sum_{i=1}^{P(P-1)+3} \frac{\left(h_j^{P^{sb}} - h_j^{G^{sb}}\right)^2}{h_j^{P^{sb}} + h_j^{G^{sb}}}, \tag{5.33}$$

where P^{sb} and G^{sb} are the sub blocks of the probe and gallery images, and $h_j^{P^{sb}}$ and $h_j^{G^{sb}}$ are the j^{th} histogram bins of the probe and gallery sub blocks, respectively. The main advantage of the local based face recognition techniques, is their capability to reduce the influence of occluded image partitions on the recognition result by assigning lower weights to the occluded regions, since they are ineffective in differentiating between facial images of different individuals, and emphasizing on the sub blocks containing discriminative information [23]. Local entropy, (5.14), was proposed in [23] which provides weights based on the information property of image sub-blocks. Local classifiers' decisions were combined and identity of the corresponding probe image was determined by adopting a decision fusion strategy. The authors of [23] utilized weighted majority voting approach, which is straightforward without learning requirement. The majority of sum of calculated local weighted votes gives the final class label.

Zhang et al. proposed LBP on the magnitude of Gabor features as the local Gabor binary pattern histogram (LGBPH) technique [39]. Representation of 2-D Gabor filter is similar to the human visual system. Gabor extracts image features at different scales and orientations, which is the multiplication of a sinusoid function by a Gaussian kernel as follows [31],

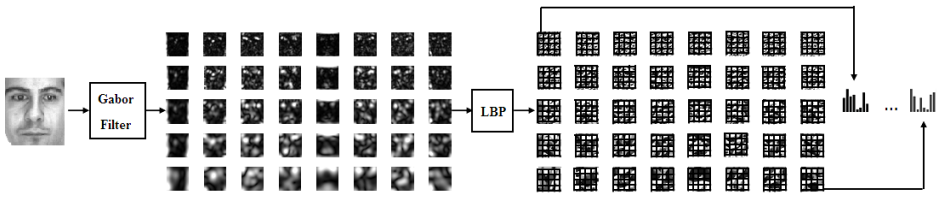

Figure 5.7. Proposed face recognition approach in [39].

$$\psi_{s,o}(x,y) = \frac{{f_s}^2}{\pi\delta\theta} \times e^{-\left[\left(\frac{{f_s}^2\hat{x}^2}{{\delta x}^2}\right)+\left(\frac{{f_s}^2\hat{y}^2}{{\delta y}^2}\right)\right]} \times e^{-(j2\pi.f_s.\hat{x})}, \tag{5.34}$$

$$\begin{cases} s = 0,1,2,\dots,S_{max}-1 \\ o = 0,1,2,\dots,O_{max}-1 \\ \hat{x} = +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \tag{5.35}$$

where S_{max} and O_{max} are the maximum number of scale and orientation, respectively. $\beta_o = (o\pi)/8$, $f_s = f_{max}/\left(\sqrt{2}\right)^s$, f_{max} is the maximum frequency of the filters, and δ_x and δ_y are the sharpness of x and y axes, respectively. In order to obtain Gabor images, the image is convolved with a filter bank of $S_{max} \times O_{max}$ filters. As seen in Fig 5.7, Gabor images are divided into small sub blocks and LBP histograms are extracted from each sub region and concatenated to make the feature vector of the image [39].

Zhang et al. also proposed another face recognition algorithm, with robustness against occlusion, using the same LGBPH technique as in [39] and employed the Kullback–Leibler divergence (KLD)-based method in order to assign weighs to the local histograms in the concatenation stage and reduce the effect of occluded image partitions by assigning lower weights to those regions [40]. KLD between the histograms of occluded sub regions and the mean histogram of gallery sub blocks is calculated as follows,

$$KLD_j = \sum_{i=1}^{N_b} H_j^P(i) \times \log\left(\frac{H_j^P(i)}{\hat{H}_j(i)}\right), \qquad j = 1,2,\dots,N_b, \tag{5.36}$$

$$\hat{H}_j = \frac{1}{N_G}\sum_{i=1}^{N_G} H_j^{G_i}, \tag{5.37}$$

Table 5.2. Accuracy of the face identification algorithms for AR database (%).

Method	Session 1		Session 2	
	sun glasses	scarf	sun glasses	scarf
Loc-Pb[33]	80	82	54	48
LGBPH[39]	80	98	62	96
KL-LGBPH[40]	84	100	80	96
AWPPZM[35]	70	72	59	60
LBP-MV[23]	100	100	100	100

where H_j^P and $H_j^{G_i}$ are histograms of j^{th} sub block in the probe and i^{th} gallery images and N_b and N_G are number of histogram bins and non-occluded gallery images, respectively. Probability of occlusion in the image sub blocks is obtained by comparing the local KLD's with a threshold [40].

Table 5.2 shows recognition accuracies of the mentioned occlusion-robust face recognition techniques on the AR database [36] with 50 random subjects and for each individual only the first sample with neutral expression in session 1 is used in the gallery set and images with sun glasses and scarves in session 1 and 2 are employed in four probe sets [23].

5.3 CONCLUSION

The performance of commercial face recognition systems is extremely degraded in real world applications such as law enforcement and security surveillance due to lighting variation, partial occlusion, facial expression, blur effect or head pose. In this chapter, a variety of state of the art face recognition strategies to deal with the appearance changes caused by illumination variation and partial occlusion are studied. Illumination insensitive face recognition methods are divided into three groups; illumination insensitive preprocessing, illumination invariant descriptors, and block-based pattern recognition. Moreover, partial occlusion corrupts some portion of discriminative facial information. Local based techniques are applied where the effect of occluded regions, on the identification performance, is excluded or reduced.

5.4 REFERENCES

[1] H. Han, S. Shan, X. Chen, W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recogn*., vol. 46, pp. 1691-1699, 2013.

[2] Y. Adini, Y. Moses, S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 19, pp. 721-732, 1997.

[3] J. R. Solar, J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different preprocessing approaches," *Pattern. Recogn. Lett.*, vol. 29, pp. 1966-1979, 2008.

[4] S. Shan, W. Gao, B. Cao, D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Proc. IEEE. Int. Wksp. Anal. Mod. (AMFG)*, 2003, pp. 157-164.

[5] S .M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, and et.al., "Adaptive histogram equalization and its variations," *Lect. Notes. Comput. Sc*., vol. 39, pp. 355-368, 1987.

[6] P. H. Lee, S. W. Wu, Y. P. Hung, "Illumination compensation using oriented local histogram equalization and its application to face recognition," *IEEE Trans. Image Process*., vol. 21, pp. 4280-4289, 2012.

[7] Z. C. Lian, M. J. Er J. Li, "A novel face recognition approach under illumination variations based on local binary pattern," in *Proc. Int. Conf. Comp. Anal. (CAIP)*, pp. 89-96, 2011.

[8] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. Image Process*., vol. 18, pp. 2599-2606, 2009.

[9] B. Xu, T. Zhang, Z. W. Shang, "Multi-scale invariant abstracted under varying illumination," in Proc. *IEEE Int. Conf. Wavelet. Anal. Pattern Recogn. (ICWAPR)*, pp. 28-32, 2012.

[10] A. K. Jain, *Fundamentals of Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[11] B. Wang, W. Li, W. Yang, Q. Liao, "Illumination normalization based on Weber's law with application to face recognition," *IEEE Signal Process. Lett*., vol. 18, pp. 462-465, 2011.

[12] M. Turk, A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci*., vol. 3, pp. 71-86, 1991.

[13] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 19, pp. 711-720, 1997.

[14] S. T. Roweis, L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000.

[15] M. Belkin, P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput*., vol. 15, pp. 1373-1396, 2003.

[16] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, "Subspace learning from image gradient orientations," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, pp. 2454-2466, 2012.

[17] H. Wang, S. Z. Li, Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *Proc. IEEE Int. Conf. Autom. FGR. (AFGR)*, 2004, pp. 819-824.

[18] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," *Pattern Recogn*., vol. 45, pp. 1299-1305, 2012.

[19] A. Baradarani, Q. M. J. Wu, M. Ahmadi, "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering," *Pattern Recogn*., vol. 46, pp. 57-72, 2013.

[20] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.

[21] W. Chen, M. J. Er, S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst. Man. Cyb*., vol. 36, pp. 458-466, 2006.

[22] T. Ahonen, A. Hadid, M. Pietikainen, "Face recognition with local binary patterns," *Lect. Notes. Comput. Sc*., vol. 3021, pp. 469-481, 2004.

[23] S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," in *International Conf. on Pattern Recognition (ICPR12)*, Tsukuba, November 2012, pp. 1699-1702.

[24] X. Tan, B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process*., vol. 19, pp. 1635-1650, 2010.

[25] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, "Illumination invariant feature extraction and mutual information based local matching for face recognition under illumination variation and occlusion," *Pattern Recogn*., vol. 44, pp. 2576-2587, 2011.

[26] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *P. Soc. Photo Opt. Ins*., vol. 2, pp. 121-167, 1998.

[27] S. Georghiades, P. N. Belhumeur, D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, pp. 643-660, 2001, The YaleB face database, Available (online) at: <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>.

[28] T. Charoenpong, "A survey of face occlusion detection for visual surveillance system," in *Proc. Int. Conf. Knowl. Smart Tech., KST*, 2010, pp. 24-25.

[29] M. Storer, M. Urschler, H. Bischof, "Occlusion detection for ICAO compliant facial photographs," in *Proc. IEEE. Comp. Soc. CVPR. Wksp. (CVPRW)*, 2010, pp. 122-129.

[30] M. Rui, A. Hadid, J. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Proc. IEEE Int. Conf. Autom. FGR. (AFGR)*, 2011, pp. 442-447.

[31] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression," *IEEE Trans. Acoust. Speech Signal. Process.*, vol. 36, pp. 1169-1179, 1988.

[32] T. Y. Kim, K. M. Lee, S. U. Lee, C. H. Yim, "Occlusion Invariant Face Recognition Using Two-Dimensional PCA," *Comm. Com. Inf. Sc*., vol. 4, pp. 305-315, 2007.

[33] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 24, pp. 748-763, 2002.

[34] H. J. Oh, K. M. Lee, S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images," *Image Vision Comput*., vol. 26, pp. 1515-1523, 2008.

[35] H. R. Kanan, K. Faez, Y. Gao, "Face recognition using adaptively weighted patch PZM array from a single exemplar image per person," Pattern Recogn., vol. 41, pp. 3799-3812, 2008.

[36] A. Martinez, R. Benavente, "The AR face database," *Technical Report CVC 24 Purdue Univ*., 1998, Available (online) at:
< http:// www2.ece.ohio-state.edu/~aliex/ARdatabase.html >.

[37] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, pp. 210-227, 2009.

[38] B. Topc, H. Erdogan, "Decision fusion for patch-based face recognition," in *Proc. IEEE Int. Conf. Pattern Recogn. (ICPR)*, 2010, pp. 1348-1351.

[39] W. Zhang, S. Shan, W. Gao, X. Chen, H Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, pp. 786-791, 2005.

[40] W. Zhang, S. Shan, X. Chen, W. Gao, "Local Gabor binary patterns based on Kullback–Leibler divergence for partially occluded face recognition," *IEEE Signal Process. Lett*., vol. 14, pp. 875-878, 2007.

Chapter 6

Human Face Recognition under Occlusion Using LBP and Entropy Weighted Voting

Human Face recognition is one of the most interesting research areas due to some challenges, such as occlusion, illumination variation, facial expression and one sample per person problem. Partial occlusion is a critical challenge in real world applications. In holistic based methods, which use the whole face area, all extracted features can be affected by appearance changes. Nevertheless, partial accessories, have impact on some parts of face. In that case, local appearance based approaches are more adequate to deal with local changes, which divide face into small patches and extract features from each region separately. A local probabilistic subspace method was used by Martinez [1] for occluded face recognition with one sample per subject. Kanan et al. in [2-4] proposed Gabor, PCA and Pseudo Zernike moment, respectively, to extract features of patches in one exemplar image per person and the identity information and likelihood of occlusion was utilized to weight the contribution of each region. Ekenel et al. [5], arranged the DCT coefficients of each 8×8 image block, using a zig-zag scan and concatenated the coefficients of different blocks after feature selection and normalization. Finally, a face alignment algorithm was proposed at classification step to deal with occlusion problem. A sparse representation was proposed by Wright et al. [6]. Therefore, the choice of features is not critical anymore but it requires many sample images. Zhang et al. in [7] proposed the LBP on magnitude of Gabor

features and concatenated the histograms of all sub regions. In [8] an improvement in the occluded face recognition performance was proposed by applying a Graphical model and Laplace equation to complete images. Furthermore Perez et al. [9], employed a combination of an entropy weighting approach on Gabor jets with Borda count voting classification to improve recognition rate of face images with occlusion.

In this chapter, we aim to recognize faces under occlusion using one sample per subject in the gallery set. LBP descriptor which is robust to rotation, scaling and illumination variation, is applied on the local blocks of image to extract texture features and a Chi-Square classifier to compare local histograms. Image entropy generates weights to combine local classifiers results and a voting scheme eventuates the final decision. We can simply achieve extremely high recognition rate compared to other approaches without computational burden.

6.1 LOCAL BINARY PATTERN DESCRIPTOR

LBP is one of the most successful feature extractors in local appearance based methods, which extracts texture features of image by comparing each pixel with its neighbors. There is no training requirement which makes it fast and easily integrable into the new data sets and also it is robust to rotation, scaling and illumination variation [10]. At each pixel position in the image, P neighbor pixels on a neighborhood with radius R, are compared with the center pixel. Therefore, we have a binary code of P bits for each pixel in the image as illustrated in Fig 6.1.

This procedure is done for each image partition and binary codes are used to make histograms for each sub-block. The code is uniform if there are at most two bitwise transitions from 0 to 1 or 1 to 0. Each histogram has $P(P-1)+2$ bins for uniform and 1 bin for all non-uniform patterns, totally $P(P-1)+3$ bins [10].

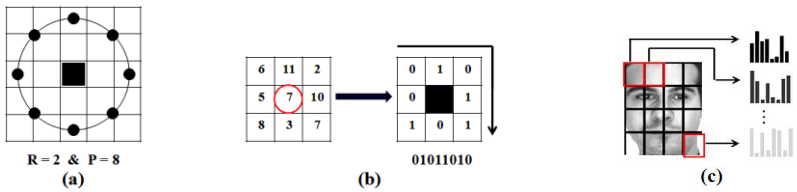

Figure 6.1. Different steps of LBP: (a) radius, R ,and number of neighboring, P, (b) comparison between each pixel and its neighbors and (c) using binary codes to make histograms for image sub blocks.

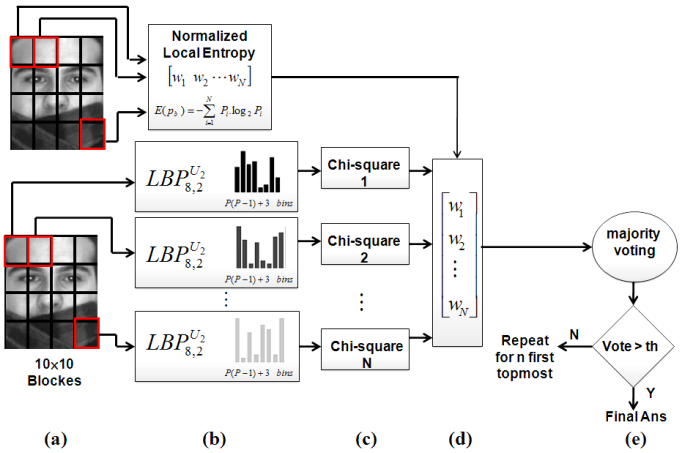

Figure 6.2. Proposed algorithm: (a) image partitioning, (b) applying LBP on each block, (c) classification of sub blocks using Chi-Square metric, (d) producing local entropy and summing the weighted votes and (e) thresholding the majority vote to find final decision.

6.2 PROPOSED BLOCK-BASED ALGORITHM

The proposed algorithm in this thesis is illustrated in Fig 6.2. As shown, it contains 6 main steps. (i) the face image is partitioned into some non-overlapping sub-blocks of size $m \times m$ pixels. (ii) LBP descriptor is applied on each block to obtain local texture features of image. (iii) local histograms are compared with the histograms of corresponding blocks in the gallery set and classified separately by utilizing Chi-Square matching method [10]. (iv) in the weighting step, we apply entropy on each sub-block to find its information content and its importance to the final decision. (v) decision fusion is done using voting scheme to find the majority vote. (vi) final answer is reclaimed by thresholding votes. In case it is lower than the threshold, classification step is repeated for the n topmost ranks.

6.3 OPTIMUM BLOCK SIZE

In most local appearance based face recognition approaches, image is divided into some non-overlapping equal-sized sub-images. The size of sub-blocks affects recognition rate of the system. Very large size adds more complexity and does not meet block-based idea. However, satisfactory information could not be provided by very small blocks. In this chapter, we proposed $m \times$

m rectangular sub-blocks. If the image size is $r \times c$ pixels, then the number of sub-blocks is equal to $N = \frac{r}{m} \times \frac{c}{m}$.

6.4 CLASSIFICATION

Features of each sub-block, obtained in previous section, are fed into classifier step to find the class assigned to each block. The best way to classify histograms is to use one of the histogram similarity measures to compare gallery and probe sub-blocks, such as histogram intersection, log-likelihood or Chi-Square statics [10], among which the latter leads to better recognition rates according to our experimental results and is as follows,

$$\mathcal{X}^2(p_b, g_b) = \sum_{j=1}^{P(P-1)+3} \frac{\left(p_{b_j} - g_{b_j}\right)^2}{p_{b_j} + g_{b_j}}, \tag{6.1}$$

where p_{b_j} and g_{b_j} are j^{th} histogram bins in the probe sub-block p_b and gallery sub-block g_b, respectively.

6.5 LOCAL ENTROPY WEIGHTING

The main advantage of the block based methods, is their capability to reduce the effect of occluded parts of image on the final decision by assigning lower weights to them, since they are not effective to differentiate faces, and emphasizing on the facial partitions with more useful information. Well-known Fisher linear discriminant (FLD) method can be applied to find the weights, which determines the discrimination power of image blocks [11], [12]. This leads to good results, but to construct the subspace, it needs a large set of database with more than one sample per person. Thus, it is not suitable to deal with one sample problem. Local image entropy which is utilized in our proposed approach, provides adequate weights according to the information property of each sub-block and is as follows,

$$E(p_b) = -\sum_{i=1}^{N_t} P_i \log_2(P_i), \tag{6.2}$$

where N_t is the overall number of gray levels in the probe image sub-block and P_i is the i^{th} gray value's probability in that block. The normalized entropy values are between 0 and 1 [14].

6.6 MAJORITY VOTING SCHEME

In order to combine the results of local classifiers and find the ultimate class of the probe image, we need to adopt a decision fusion method, which gives much better results compared to feature fusion [12]. Weighted sum rule and majority voting scheme are two fixed rule combination approaches which are straightforward without any training requirement [13]. In this chapter we proposed a weighted majority voting scheme, in which according to the local entropy weights provided in the previous section, we sum the weights of votes for each class. The final decision is the label of the majority vote.

To achieve better recognition accuracy, the maximum vote is compared with a threshold, which is achieved by experimental trial. In case it is less than the threshold, the classification and majority voting steps are repeated for the n topmost number of votes. This does not burden considerable computational complexity on the system due to the fact that n is much less than the number of classes [14].

6.7 EXPERIMENTAL RESULTS

In order to appraise the performance of the proposed algorithm, two experiments are conducted on the AR face database. AR database contains more than 3000 facial images of 136 people (76 men and 60 women) [15]. 26 images of each subject were taken in two sessions with 2-week time gap (13 images per session). The database consists of images with occlusion (due to scarf or sun glasses), facial expression (neutral, smile, scream and anger) and illumination variation (left, right and all side light). Therefore, it is feasible to evaluate our proposed approach with all challenges. We use the grey images of size 160×120 with intensity adjustment in MATLAB, to reduce the effect of illumination variation. We conduct two different experiments and compare the results with some well-known existing algorithms.

6.7.1 Experiment 1

In this experiment, 50 random selected subjects are partitioned into 10×10 pixel sub-blocks. To assess the single sample problem, only the first non-occluded images with neutral expression in the first session are employed for all subjects in the gallery set. The first occluded images with scarf and sun glasses in both sessions 1 and 2 are utilized in four probe sets. The LBP histogram has 59 bins (P=8 and R=2) and in the last stage of voting scheme n is equal to 6 topmost ranks. Table 6.1 shows the recognition results in comparison with some recent methods.

As can be seen in Table 6.1, our proposed method could achieve 100% recognition in all cases which significantly outperforms other methods, with less complexity compared to [2] and [7].

We also repeated the same experiment for 100 subjects (50 men and 50 women) and showed the recognition rates in Table 6.1. In this case, we have a slight reduction in the recognition results compared to the 50-subject experiment. However, our algorithm outperforms other methods.

6.7.2 Experiment 2

In this experiment, 100 subjects (50 men and 50 women) were employed. To evaluate our algorithm in the case that different challenges exist such as occlusion, expression and illumination variation, all 7 non-occluded images in session 1 are taken as gallery set. Three probe sets consist

Table 6.1. Recognition rate percentages of experiment 1.

Method	Session1		Session2		Numb of class	Image Size
	Scarf	Sun	Scarf	Sun		
Proposed method	100	100	100	100	50	160×120
Proposed method	98	98	94	97	100	160×120
LocPb[1] *	82	80	48	54	50	-
AWPASG[2]*	84	38	70	20	50	160×160
AWPCA[3]*	76	44	58	22	50	160×160
AWPPZM[4]*	72	70	60	59	50	160×160
LGBPHS[7]*	98	80	96	62	50	-
LA-DCT[5]*	97.3	98.2	93.6	95.5	110	-

∗ Results extracted from original source

Table 6.2. Recognition rate percentages of experiment 2.

Method	Scarf	Sun	Non-occluded	Numb of class	Image Size
Proposed Method	98.17	99	99.57	100	160×120
SRC[6]*	93.50	97.50	95.70	100	165×120
GL PCA+GW[8]*	60.90	76.60	-	121	64×64
LMGEW-BTH[9]*	99.50	96	99.29	100	-

∗ Results extracted from original source

of 7 non-occluded images in session 2 and 6 occluded images with scarf and 6 faces with sun glasses in sessions 1 and 2, respectively. All other conditions are remained the same as experiment 1. Table 6.2 shows the results in comparison with some of the well-known methods. According to the table, our approach outperforms all other methods, except just in the case of scarf occlusion that LMGEW-BTH [9] performs better. However, that approach, which is a combination of entropy weighted Gabor jets and a modified Borda count scoring classification, compared to our method (only a simple LBP and voting scheme) burden much more complexity in both feature extraction and classification steps to the system and it requires very precise eye localization.

6.8 CONCLUSION

A novel local block-based algorithm is proposed in this chapter which benefits from the local binary pattern (LBP) texture feature descriptor, and decision fusion by means of entropy weighted majority voting approach. This algorithm leads to satisfactory stability against partial occlusion, when only one sample per class is attainable. By applying local entropy to find the information content of each sub-block, we can reduce the contribution of some partitions of the probe image with undesirable appearance changes at final decision, by devoting lower weights to them in the voting scheme. We have achieved extremely high recognition rates on very challenging AR face database, in comparison with recent well-known methods, without imposing computational complexity.

6.9 REFERENCES

[1] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, (6) pp. 748–763, June 2002.

[2] H. R. Kanan, K.,Faez, "Recognizing faces using adaptively weighted sub-Gabor array from a single sample per enrolled subject," *Pattern Recognit*., vol. 28, (3) pp. 438-448, March 2010.

[3] H. R. Kanan, M. S. Moin "Recognizing partially occluded faces from a single exemplar image per person," *Lecture Notes in Computer Science*, vol. 5576, pp. 100-109, 2009.

[4] H. R. Kanan, K. Faez, Y. Gao, "Face recognition using adaptively weighted patch PZM array from a single exemplar image per person," *Pattern Recognition*, vol. 41, (12) pp. 3799–3812, December 2008.

[5] H. K. Ekenel, R. Stiefelhagen, "Why is facial occlusion a challenging problem," *Advances in Biometrics. Springer Berlin Heidelberg*, 2009, pp. 229–308.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, (2) pp. 210-227, February 2009.

[7] W. Zhang, S. Shan, W. Gao, X, Chen, H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," *ICCV*., 2005, pp. 786-791.

[8] Y. Deng, Q. Dai, Z. Zhang, "Graph Laplace for occluded face completion and recognition," *IEEE Trans. Image Process*., vol. 20, (8) pp. 2329-2338, August 2011.

[9] C. A. Perez, L. A. Cament, L. E. Castillo, "Local matching Gabor entropy weighted face recognition," *ICIG*, August 2011, pp. 179-184.

[10] T. Ahonen, A. Hadid, M. Pietikainen, "Face recognition with local binary patterns," *ECCV*, 2004, pp. 469-481.

[11] Y. Su, S. Shan, X. Chen, W. Gao, "Hierarchical ensemble of global and local classifiers for Face recognition," *IEEE Trans. Image Process*., vol. 18, (8) pp.1-8, August 2009.

[12] B. Topc¸, H. Erdogan "Decision fusion for patch-based face recognition," *ICPR*, August 2010, pp. 1348-1351.

[13] F. Roli1, J. Kittler, G. Fumera, D. Muntoni, "An experimental comparison of classifier fusion rules for multimodal personal identity verification systems," *MCS*, 2002, pp. 325–335.

[14] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi "Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion," *Pattern Recognition*., vol. 41, (10-11) pp. 2576-2587, October 2011.

[15] A. Martinez R. Benavente, "The AR face database," *CVC Tech. Rep*., (24), June 1998.

Chapter 7

A Local Gradient-Based Illumination Invariant Face Recognition Using LPQ and Multi-Resolution LBP Fusion

Face recognition is one of the most important biometric identification methods which is used for many applications such as law enforcement, video surveillance and human-computer interaction. In spite of improvement in identification accuracies, it is still one of the most challenging areas in the field of pattern recognition due to variations in illumination, pose, facial expression, blur and occlusion. Illumination is an effective factor which might degrade the performance of recognition system significantly. Within the past few years different strategies have been employed to overcome the illumination variation effect. In [1] an orientated local histogram equalization (OLHE) is proposed, which captures an illumination insensitive face representation while preserving rich information on the edge orientations. Gradient-based preprocessing techniques find the illumination insensitive representation of the image using the fact that the illumination component changes slowly, in comparison with abrupt changes of reflection [2]. In [3] the illumination insensitive representation is derived using the ratio of directional gradients. In [4] the authors also improved their technique by wavelet filtering of partial gradients to obtain the multi-resolution illumination invariant analysis. In [5] a subspace

learning method is proposed using the gradient orientation and cosine-based correlation which leads to an illumination and occlusion invariant image representation. In [6] the illumination insensitive image is calculated using Weber law which is the tangent inverse of the ratio between the local intensity variations, using Laplace operator.

Another category of illumination insensitive face recognition techniques are based on the estimation of reflectance component [2]. In the proposed algorithm in [7] the illumination enhanced image is decomposed into frequency sub-bands using a double-density dual-tree complex wavelet transform (DD-DTCWT). The image is reconstructed based on the low frequency and thresholded high frequency coefficients. Maximum filtering and edges amplification give the illumination insensitive image. Discrete cosine transform (DCT) is applied on the image logarithm in [8] and by eliminating the low frequency coefficients, the illumination variation is compensated. In the method proposed in [9] a combination of Gamma correction, difference of Gaussian (DoG) filtering, masking and contrast equalization is used along with a kernelized illumination robust local ternary pattern (LTP).

Rather than holistic-based methods, local-based techniques suffer much less from local appearance changes. Moreover, holistic methods do not tackle the asymmetry in the intensity due to the shadow which is caused by side-lighting [2]. In [10] luminance component is extracted and removed by applying the maximum filter on the logarithmic illumination-reflection model and majority voting combines the recognition result of image partitions.

In this chapter, a local based algorithm which is the combination of image preprocessing and fusion of two illumination invariant local descriptors is proposed as follows: i) an illumination insensitive image representation is achieved based on the ratio of the gradient magnitude to the original image intensity, ii) the preprocessed image is divided into small sub blocks, iii) local phase quantization (LPQ) and multi-scale LBP extract discriminant characteristics of sub images, iv) local nearest neighbor classifiers based on the Chi-Square distance metric[11] acquire the distance scores between the sub blocks in probe and gallery images, v) score-level fusion is utilized to combine the local distances to obtain the most probable match for the probe image, vi) decision-level fusion is employed to combine the results of two matching techniques. Entropy, class posterior probability and mutual information are employed as adequate weights of fusion components. The block diagram of the algorithm is illustrated in Fig. 7.1.

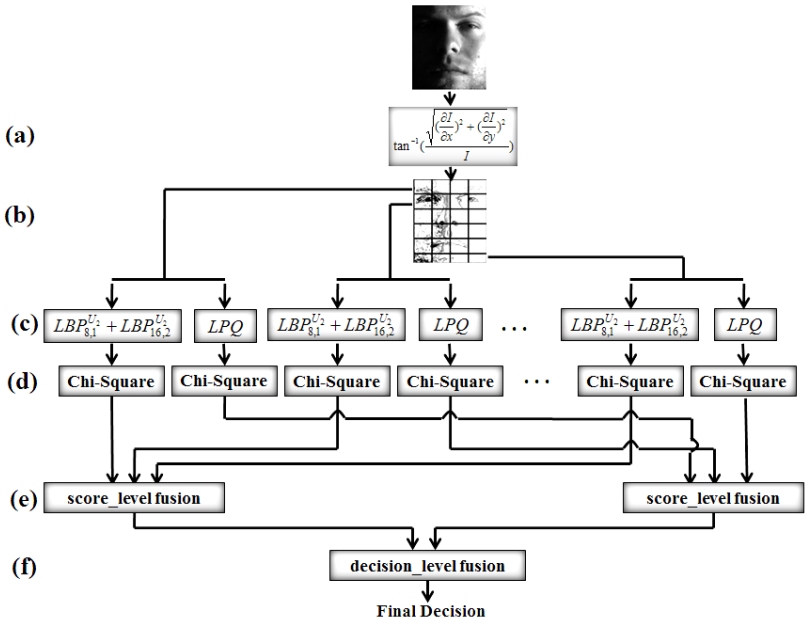

Figure 7.1. Block diagram of the proposed algorithm: a) Illumination normalization, b) Image partitioning, c) LBP and LPQ descriptors for image partitions, d) Chi-Square based local nearest neighbor classifiers, e) score-level fusion to combine local distance measures and f) decision-level fusion to combine the results of two matching techniques.

The rest of chapter is organized as follows. In section 7.1 and 7.2, illumination insensitive representation and image partitioning are described. Section 7.3 and 7.4 are explaining the LBP and LPQ descriptors and the classification approach in details. In section 7.5 and 7.6 the score and decision level fusion techniques and their devoted weights are described mathematically. Section 7.7 is related to the implementation parameters, face databases and simulation results. The chapter is concluded in section 7.8.

7.1 ILLUMINATION EFFECT SUPPRESSION

To minimize the influence of the illumination variation on face recognition, different methods of image normalization can be applied. In this chapter, we have proposed an image representation based on the gradient domain processing. In the gradient domain, image representation is more discriminative and dependency between pixels is taken into consideration [3]. The proposed

method is applied directly on the image without training and reconstruction requirement. Detailed procedure of preprocessing method is as follows.

The general form of the Lambertian reflection model of image is as follows,

$$i(x,y) = \rho(x,y).L(x,y), \tag{7.1}$$

where $i(x,y)$ is the image pixel value at point (x,y), and $\rho(x,y)$ and $L(x,y)$ are the reflectance and luminance values at the same image location. We intend to find the illumination insensitive representation of the image. Due to the fact that $L(x,y)$ is the low frequency component of the image and varies slowly, despite the abrupt changes in the reflection component, it is generally considered to be smooth, which means that $L(x+\Delta x,y) \approx L(x,y)$ and $L(x,y+\Delta y) \approx L(x,y)$. In the gradient domain, the ratio of image derivative along the y direction to the derivative of image along the x axis is proven to be insensitive to illumination (the proof is in [3]),

$$i(x+\Delta x,y) \approx \rho(x+\Delta x,y).L(x,y). \tag{7.2}$$

$$i(x,y+\Delta y) \approx \rho(x,y+\Delta y).L(x,y). \tag{7.3}$$

Subtracting (2) and (3) from (1) and dividing by Δx and Δy, respectively, we have the image partial derivatives,

$$\frac{i(x+\Delta x,y)-i(x,y)}{\Delta x} \approx \frac{[\rho(x+\Delta x,y)-\rho(x,y)]}{\Delta x}.L(x,y) \Rightarrow \quad \frac{\partial i(x,y)}{\partial x} \approx \frac{\partial \rho(x,y)}{\partial x}.L(x,y). \tag{7.4}$$

$$\frac{i(x,y+\Delta y)-i(x,y)}{\Delta y} \approx \frac{[\rho(x,y+\Delta y)-\rho(x,y)]}{\Delta y}.L(x,y) \Rightarrow \quad \frac{\partial i(x,y)}{\partial y} \approx \frac{\partial \rho(x,y)}{\partial y}.L(x,y). \tag{7.5}$$

The magnitude of the image gradient is as follows,

$$GA = \sqrt{\left(\frac{\partial i(x,y)}{\partial x}\right)^2 + \left(\frac{\partial i(x,y)}{\partial y}\right)^2} \approx \sqrt{\left(\frac{\partial \rho(x,y)}{\partial x}.L(x,y)\right)^2 + \left(\frac{\partial \rho(x,y)}{\partial y}.L(x,y)\right)^2}$$

$$= L(x,y).\sqrt{\left(\frac{\partial\rho(x,y)}{\partial x}\right)^2 + \left(\frac{\partial\rho(x,y)}{\partial y}\right)^2}. \tag{7.6}$$

Consequently, the ratio of the gradient magnitude to the original image intensity value suppresses the illumination effect significantly and leads to a more robust illumination insensitive image representation with more discrimination power compared to the Gradientface in [3], as shown in Fig 7.2,

$$\frac{GA}{i} = \frac{L(x,y).\sqrt{\left(\frac{\partial\rho(x,y)}{\partial x}\right)^2+\left(\frac{\partial\rho(x,y)}{\partial y}\right)^2}}{L(x,y).\rho(x,y)} = \frac{\sqrt{\left(\frac{\partial\rho(x,y)}{\partial x}\right)^2+\left(\frac{\partial\rho(x,y)}{\partial y}\right)^2}}{\rho(x,y)}. \tag{7.7}$$

In order to avoid the ambiguity due to the zero values of image intensity in the above ratio and reduce the noise level [6], we apply the tan^{-1} of the ratio, as follows,

$$\delta = tan^{-1}\left[\frac{GA}{i}\right], \quad \text{where} \qquad 0 \le \delta < 2\pi. \tag{7.8}$$

However, we need to smoothen the image, using a Gaussian filter to make the calculation of the gradient operation more stable against noise [3]. The standard deviation of Gaussian filter,σ, is chosen experimentally through an exhaustive search. $\sigma = 0.75$ pixel is found to be the best

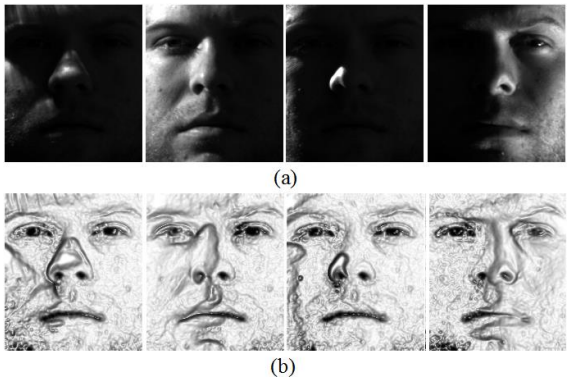

Figure 7.2. Illumination invariant representation of sample images of Yale B database: (a) original images and (b) preprocessed images using proposed method.

value. Figure 7.2 illustrates some samples of the illumination insensitive images in the Yale B database using the proposed method.

7.2 IMAGE PARTITIONING

Instead of applying the face recognition algorithm on the whole image, we have proposed a local appearance based method in which the preprocessed image is divided into small sub blocks and the entire process is performed on each sub image and the results are fused together to find the final decision. The image partitioning can be performed in the form of equal or unequal and overlapped or non-overlapped divisions. Our exhaustive search demonstrated that equal-size rectangular sub-blocks without overlapping would yield better recognition results. Very large block size adds complexity and does not meet block-based idea. However, satisfactory information could not be provided by very small blocks [12]. In this work, we proposed $m \times m$ rectangular sub blocks. If the image size is $r \times c$ pixels, then the number of sub blocks is equal to $\frac{r}{m} \times \frac{c}{m}$ [12].

7.3 FEATURE EXTRACTION

It is proven that combining feature extractors can result in better performance than employing one descriptor. Among local descriptors with significant performance, local binary pattern and local phase quantization are known to be computationally fast with simple calculations which are insensitive against monotonic image degradation [11, 13].

7.3.1 Multi Resolution Uniform LBP

LBP is one of the most successful local feature extractors, which extracts texture features of the image by comparing each pixel with its neighbors in a small neighborhood [12]. There is no training requirement which makes it fast and easy to integrate into the new data sets [11]. Furthermore, due to the application of histograms as the feature sets, it is robust against rotation and scaling (caused by errors in registration). Also, the image-size dimension can be reduced to

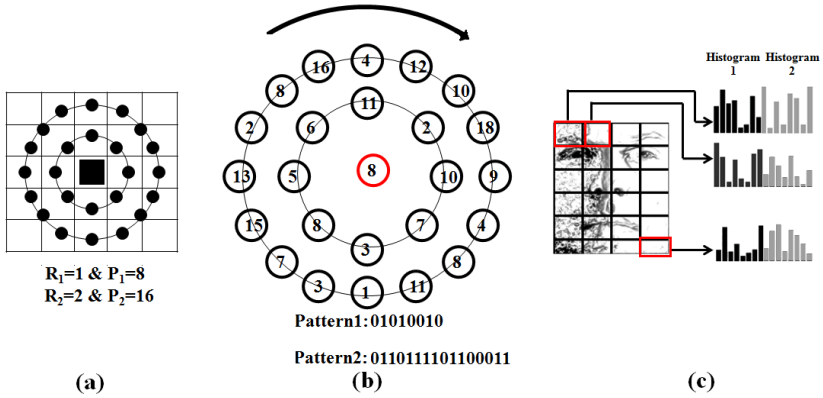

Figure 7.3. Different steps of LBP: (a) radius of two neighboring windows, R_1& R_2, and number of neighbors , P_1& P_2, (b) binary patterns obtained by comparison between the center pixel and its neighbors and (c) utilizing binary codes to make the histograms and concatenate them for each image sub-block.

the number of histogram bins [11]. Since the gray values of neighbors are compared, it is robust against monotonic variations in the face image. Thus, LBP is one of the illumination insensitive descriptors [11]. *R* is the radius of neighborhood and *P* is the number of neighbors which are compared with the center pixel [12]. Therefore, we have a *P*-bit binary pattern for each pixel as illustrated in Fig 7.3.

This procedure is carried out for each image sub-block and binary patterns are used to make the histograms for each of them. In this work we ignore the non-uniform patterns due to the fact that considerable amount of discriminating information is preserved by taking only uniform patterns into consideration [11]. The code is uniform if there are at most two bitwise transitions from 0 to 1 or 1 to 0. Each histogram has $P(P-1)+2$ bins for uniform and 1 bin for all non-uniform patterns, totally $P(P-1)+3$ bins. Therefore, the calculation complexity is also reduced [11].

In order to obtain a multi-resolution representation, we repeat the above procedure for two different radiuses and concatenate the feature vectors extracted for each neighborhood as shown in Fig 7.3 In this chapter we concatenate the histograms of $LBP^{uni}_{R_1,P_1}$ and $LBP^{uni}_{R_2,P_2}$, where "uni" stands for uniform pattern. Although the image description has the problem of high dimensionality, $(P_1(P_1-1)+3+P_2(P_2-1)+3$ bins per sub-block), the face representation would be much more accurate with more discrimination power and robust against scale variation.

7.3.2 LPQ

Local phase quantization (LPQ) is another widespread histogram-based feature extractor from the family of local texture descriptors which performs the assessment of phase in a local window at the pixel position [13]. Local phase analysis in frequency domain leads to a detailed illumination insensitive texture description of face images [14, 15]. LPQ is also insensitive against another image degradation, blur effect, which happens usually in real world applications such as video surveillance, which is caused by out of focus of camera or object motion. LPQ is a common illumination and blur insensitive feature extractor [16, 17]. The blurred image can be modeled as follows,

$$b(x,y) = i(x,y) * p_s(x,y), \tag{7.9}$$

where $b(x,y)$, $i(x,y)$ and $p_s(x,y)$ are blurred image, original image intensity and the point spread function (PSF) of blur, respectively, and * denotes the 2D convolution [16].

The model in the Fourier domain is as follows,

$$B(u,v) = I(u,v).P_s(u,v), \tag{7.10}$$

where $B(u,v)$, $I(u,v)$ and $P_s(u,v)$ are discrete Fourier transform (DFT) of $b(x,y)$, $i(x,y)$ and $p_s(x,y)$, respectively. PSF for an ideal motion or out of focus blur, can be modeled as a rectangular which is centrally symmetric and thus the frequency spectrum is a Sinc function, which is positive at small frequencies bellow the first zero crossing. Also, in the case of atmospheric turbulence blur, the PSF is centrally symmetric and modeled by a Gaussian function and its frequency spectrum is a Gaussian which is positive at all frequencies [18, 19]. Therefore, at low frequencies, within the bandwidth of PSF, in the Fourier domain, the blur component is centrally symmetric and $P_s(u,v)$ is positive. Thus, local analysis of phase information at low frequencies leads to a blur insensitive face description [18, 19],

$$< B(u,v) =< I(u,v). \tag{7.11}$$

To have a more precise and stable local analysis of phase, we calculate the short time Fourier transform (STFT), at four specific frequencies $[u_1, v_1] = [0, a], [u_2, v_2] = [a, a], [u_3, v_3] = [a, 0]$ and $[u_4, v_4] = [a, -a]$, in a $\omega \times \omega$ window, $W(x, y)$, at each pixel position, where $a = 1/\omega$ is sufficiently small, below the first zero crossing [18],

$$B(\alpha, \beta, u, v) = \sum_{y \in W} \sum_{x \in W} b(x, y). W^*(x - \alpha, y - \beta). e^{-j2\pi(ux+vy)}, \quad (7.12)$$

where α and β are the spatial position and u and v are the spatial frequency of the window, $W(x, y)$, respectively [20].

We employ a binary quantization to encode the sign of real and imaginary parts of these four frequency coefficients as follows,

$$q_t = \begin{cases} 1 & if\ c_t \geq 0 \\ 0 & if\ c_t < 0 \end{cases}, \quad (7.13)$$

where c_t is the t^{th} element of the following 8-component vector,

$$V = [B_R(u_1, v_1), B_R(u_2, v_2), B_R(u_3, v_3), B_R(u_4, v_4), B_I(u_1, v_1), B_I(u_2, v_2), B_I(u_3, v_3), B_I(u_4, v_4)], \quad (7.14)$$

where $B_R(u_t, v_t) = Real\{B(u_t, v_t)\}$ and $B_I(u_t, v_t) = Imaginary\{B(u_t, v_t)\}$. An 8-bit pattern is assigned to an image pixel, thus for each image sub-block we have a histogram of 2^8 bins. Figure 7.4 shows the whole procedure.

7.4 LOCAL CLASSIFICATION

Features of each sub-block, which are extracted by two different descriptors, are fed into the nearest neighbor local classifiers separately. The most accurate way to classify the histograms extracted by LBP and LPQ is to utilize one of the histogram similarity measurement methods, mentioned in [11], among which Chi-Square statistics leads to better performance regarding the experimental results,

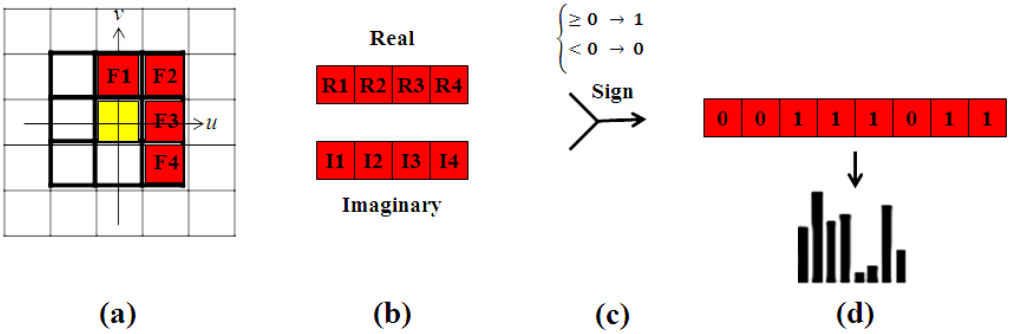

Figure 7.4. LPQ procedure: (a) STFT at four specific frequencies inside a small window at the pixel position, (b) finding the real and imaginary of the four frequency coefficients, (c) binary quantization based on the sign of real and imaginary values and (d) constructing a histogram based on the 8-bit stream.

$$D_{Chi}\left(Pb^i, Gb^i\right) = \sum_{j=1}^{N_B} \frac{(Pb^i_{B_j} - Gb^i_{B_j})^2}{Pb^i_{B_j} + Gb^i_{B_j}}, \tag{7.15}$$

where N_B is the number of histogram bins and $Pb^i_{B_j}$ and $Gb^i_{B_j}$ are the j^{th} bin in the histogram of the i^{th} sub-block of the probe image, Pb^i, and gallery image, Gb^i, respectively. A fusion technique is applied to combine these local results in order to find the distance between two images, which will be explained in the following section.

7.5 SCORE LEVEL FUSION

7.5.1 Block Distance Combination

Using the defined local classifiers, we need to combine the distance measures between the probe and gallery sub-regions to find the dissimilarity measurement of two images. In this chapter, we applied a weighted sum rule as the combination technique on the classifiers separately, as follows,

$$D(P, G) = \sum_{i=1}^{N_b} [W_{Pb^i} . D\left(Pb^i, Gb^i\right)], \tag{7.16}$$

where W_{Pb^i} is the weight which is assigned to the i^{th} sub-block of probe image and is the multiplication of two different weights as follows,

$$W_{Pb^i} = E(Pb^i).Post(C_l|Pb^i), \tag{7.17}$$

where $E(Pb^i)$ and $Post(C_l|Pb^i)$ are local entropy and logistic posteriori probability which are explained in the following sub sections.

7.5.2 Local Entropy

One of the main advantages of the local appearance based methods is their capability to reduce the effect of deformed parts of image on the recognition performance by assigning lower weights to those sub regions since they are not effective to differentiate faces [12]. Entropy is utilized in our proposed approach to find the adequate local weights, which determine the discrimination power of image sub blocks according to their information property [10, 12],

$$E(Pb^i) = -\sum_{j=1}^{N_I}[p_j.log_2(p_j)], \tag{7.18}$$

where $E(Pb^i)$ is the entropy of the i^{th} sub-block in the probe image and p_j and N_I are the probability of the j^{th} intensity value in the image sub-block and the total number of sub-block intensity levels, respectively. The local entropy is normalized to a value in the [0, 1] interval.

7.5.3 Logistic Posterior Probability

In this chapter, we have employed the block distances to find the conditional class posteriori probability of each image sub-block according to its similarity score as the degree of class support [21]. In order to calculate the class posteriori probability appropriately, we utilize the logistic link function (sigmoid function) [22]. We need to find the distance of a sub-block in the probe image to the nearest gallery sub-block per class, ${D_l}^{min}(Pb^i, Gb^i)$, which is the minimum distance of the i^{th} sub-block in the probe image to the i^{th} sub-block in the gallery samples of

l^{th} class of subjects, where $l = 1, 2, \dots, N_c$ (N_c is the total number of classes). Therefore, we have a vector of minimum distances for the i^{th} sub-block of the probe image as follows,

$$D^{min}(Pb^i, Gb^i) = \left[{D_1}^{min}(Pb^i, Gb^i), {D_2}^{min}(Pb^i, Gb^i), \dots, {D_{N_c}}^{min}(Pb^i, Gb^i)\right]. \quad (7.19)$$

The conditional posteriori probability of the i^{th} sub-block of the probe image given the l^{th} class is as follows,

$$Post(C_l|Pb^i) = sigmoid\left(\frac{\sum_{\substack{j=1 \\ j\neq l}}^{N_c} {D_j}^{min}(Pb^i)]}{{D_l}^{min}(Pb^i)}\right), \text{where} \quad sigmoid(x) = \frac{1}{1+exp^{-x}}. \quad (7.20)$$

The calculated logistic posteriori probability is normalized to a value between 0 and 1.

7.6 DECISION LEVEL FUSION

7.6.1 Classification Result

The classifier result is a decision on the class to which the probe image belongs. Again, as in section 7.3, we find the distance vector of nearest gallery image to the probe image per class,

$$D^{min}(P, G) = \left[{D_1}^{min}(P, G), {D_2}^{min}(P, G), \dots, {D_{N_c}}^{min}(P, G)\right], \quad (7.21)$$

where ${D_l}^{min}(P, G)$ is the minimum distance of the probe image to the gallery samples of l^{th} class and $l = 1, 2, \dots, N_c$.

Thus, the best match for the probe image is the class with the least distance. In order to improve the accuracy of the classification, a weighted version of distance vector is utilized as follows,

$$R_s = argmin_l\{W.D^{min}(P, G)\}, \quad \text{where} \quad \begin{cases} R_s \in [1,2, \dots, N_C] \\ l \in [1,2, \dots, N_C], \\ s \in [1,2, \dots, N_S] \end{cases} \quad (7.22)$$

$$W = [W_1, W_2, \dots, W_{N_C}], \tag{7.23}$$

where R_s is the classification result of the s^{th} matching technique and W_l is the weight which is assigned to the l^{th} class. N_S and N_C are the number of matching techniques, and the total number of classes, respectively. Mutual information is employed as the weight.

7.6.2 Mutual Information

Mutual information of two random variables A and B, which is a resemblance criterion, is as follows,

$$\mu I(A,B) = \sum_{a\in A}\sum_{b\in B}\left[P(a,b).log_2\left(\frac{P(a,b)}{P(a).P(b)}\right)\right], \tag{7.24}$$

where $P(a,b)$, $P(a)$ and $P(b)$ are the joint and marginal probability distribution functions of A and B, respectively.

Entropy, which measures the information property of random variable, can be employed to express the mutual information [10],

$$\mu I(A,B) = E(A,B) - E(A|B) - E(B|A), \tag{7.25}$$

where $E(A,B)$, $E(A|B)$ and $E(B|A)$ are joint and conditional entropy of A and B, respectively.

$$E(A,B) = -\sum_{a\in A}\sum_{b\in B}[P(a,b).log_2 P(a,b)]. \tag{7.26}$$

$$E(A|B) = -\sum_{a\in A}\sum_{b\in B}[P(a|b).log_2 P(a|b)]. \tag{7.27}$$

For 2D images, mutual information shows how identical two images are, and the more similar they are, the larger is the value of mutual information and the smaller is the joint entropy value of those two images. In this work, we find the mutual information of each probe image and the mean gallery image per class, $\mu I_l\left(P, M_G^{c_l}\right)$,

$$W = \left[(1-\mu I_1), (1-\mu I_2), \dots, \left(1-\mu I_{N_C}\right)\right], \tag{7.28}$$

where $(1-\mu I_l)$ shows the dissimilarity between the probe image, P, and $M_G^{c_l}$, the average of all gallery samples belong to the l^{th} class. The value of mutual weight is normalized and between 0 and 1.

7.6.3 Classifier Fusion

We have two classification results per probe image, which might agree or disagree on a specific class. However, we require combining these results to find the best suited decision for each probe image. In this chapter, we compare the class posterior probability of classification results, for the particular class label which they attributed to the probe image, P, using the distance measurements in (7.21),

$$Post(R_s|P) = sigmoid\left(\frac{\sum_{\substack{j=1 \\ j\neq R_S}}^{N_C} {D_j}^{min}(P,G)]}{{D_{R_S}}^{min}(P,G)}\right), \tag{7.29}$$

$$R = argmax_{R_s}\{Post(R_s|P)\} \quad R \in [1,2,\dots,N_C]. \tag{7.30}$$

7.7 SIMULATION RESULTS

7.7.1 Face Database and Simulation Settings

In order to test the stability of the proposed algorithm against image degradation caused by illumination variation, we conducted the experiments on five widely used face data bases, Yale B, Extended Yale B, AR, Multi-PIE and Face Recognition Grand Challenge (FRGC). It is possible to evaluate the algorithm robustness against sever illumination changes in uncontrolled conditions. The simulation parameters which lead to the best performance of the algorithm through the exhaustive search, are as follows: For LBP and LPQ, $R_1,\ P_1,\ R_2,\ P_2 = 1, 8, 2, 16$ and $\omega = 3$, , respectively, and the image block size is 12×12 pixels. We compared our results with some state-of-the-art approaches in this regard.

7.7.1.1 Yale B and Extended Yale B databases

The Yale B database [23] contains 5760 images of size 192×168 pixels for 10 subjects with 9 poses and 64 illumination conditions per pose. The Extended YaleB database is the expanded version from 10 subjects to 38 subjects [24], with 21,888 single light images under the same viewing conditions in Yale B. The frontal pose images (64 images per individual) are utilized in this chapter to evaluate the illumination insensitivity of the proposed algorithm. In order to compromise the comparisons with the well-known techniques in the literature, the Yale B and Extended Yale B databases are divided into five subsets corresponding to the angle between the light source and the camera axis, α. Subset 1 consists of 7 images per subject, ($\alpha < 12^{\circ}$). Subsets 2 to 5 contain 12, 12, 14 and 19 images per subject, with ($20^{\circ} < \alpha < 25^{\circ}$), ($35^{\circ} < \alpha < 50^{\circ}$), ($60^{\circ} < \alpha < 77^{\circ}$) and ($\alpha > 78^{\circ}$), respectively. Thus, for Yale B database, subsets 1 to 5 contain 70, 120, 120, 140 and 190 images and the Extended Yale B consists of 226, 454, 454, 532 and 722 faces, respectively. Fig 7.5a shows the images of the first individual in subset 5 and their illumination insensitive representation obtained by utilizing the proposed method in section 7.1.

Tables 7.1 and 7.2 show the simulation results of the first experiment on the Yale B and Extended Yale B databases. Subset 1 is used as the gallery set and subsets 2-5, are employed as four probe sets [25]. The face images are eye aligned and cropped into 192×168 gray level resolution. The results show the improved performance and effectiveness of the proposed algorithm under severe illumination changes. The precision values for subsets 4 and 5 in Table 7.2 are 99.85% and 99.32%, and recall values are 99.81% and 99.17%, respectively. In the case of subset 4 in Table 7.2, the algorithm performs significantly better than other techniques in the literature. The result of [7] shows 7 mismatches for subset 4 while the proposed method has only one misclassification.

In the second experiment on the Extended Yale B, we randomly choose five images per individual as the gallery and the remaining images as probe set. Due to random selection of the gallery samples, we repeat the experiment 50 times and compare the average result with other techniques [5]. We have also calculated the precision and recall which are 99.20% and 98.30%, respectively. As shown in Tables 7.3, the proposed algorithm outperforms the method in [5], which is a subspace learning technique using the gradient orientation, without learning requirement.

Table 7.1. Accuracy percentage of the face identification algorithms for subsets 2-5 of the Yale B database (%).

Method	Subset2	Subset3	Subset4	Subset5
Weberface[6]	100	100	96.4	96.8
LBP[26]	100	97.60	65.20	44.40
LTV[26]	100	100	98.57	100
LBP-MV[12]	100	100	95	82
MGI[4]	100	100	100	100
MPD-DCT[27]	100	100	92.57	95.95
Gradientfaces[3]	100	100	99.28	100
II+PCA [10]	100	100	98.60	98.90
II+PCA+DVS[10]	100	100	99.29	99.47
Il+PCA+WVS[10]	100	100	99.29	100
MFSR+PCA [28]	100	100	100	100
DD-DTCWT[7]	100	100	100	100
Proposed Method	100	100	100	100

Table 7.2. Accuracy percentage of the face identification algorithm for subsets 2-5 of the Extended Yale B database (%).

Method	Subset2	Subset3	Subset4	Subset5
LRC[29]	100	100	83.27	33.61
LBP-MV[12]	100	100	89.66	84.90
II+PCA+SVM[10]	100	99.78	95.44	94.68
II+PCA+1NN[10]	100	100	96.01	92.44
II+PCA+DVS[10]	100	100	96.60	95.40
Il+PCA+WVS[10]	100	100	97.91	96.54
DD-DTCWT[7]	100	100	98.68	99.03
Proposed Method	100	100	99.81	99.17

Table 7.3. Average accuracy percentage of 50 experiment on the Extended Yale B database (%).

Method	LBP [26]	Gabor-PCA [30]	Gabor-FLD [31]	IGO-PCA [5]	IGO-LDA [5]	Proposed Method
Recognition Accuracy (%)	80	73.90	64.3	95.65	97.80	98.30

Table 7.4. Accuracy percentage of the face identification algorithm on the AR database (%).

Method	SRC [25]	WSRC [33]	D-HLDO [34]	DFR [35]	GRRC-L2 [36]	CRC-RLS [37]	LBP-MV [12]	Proposed Method
Recognition Accuracy(%)	94.7	94.43	93	93.7	97.3	93.7	99	99

7.7.1.2 AR database

AR face database contains 3536 facial images of 136 people (76 men and 60 women) [32]. 26 images per subject were taken in two sessions in two weeks (13 images per session). The database consists of eye aligned images of size 168×120 pixels with illumination variation (left, right and all side light), facial expression (neutral, smile, scream and anger) and partial occlusion (scarf or sun glasses) [32]. The original and preprocessed images of session1 for the first subject are shown in Fig 7.5b. The result on AR database with medium lighting variation, facial expression and changes in the appearance within two weeks is given in Table 7.4. In this experiment we take 7 non-occluded images with different facial expressions in session 1 as the gallery images and 7 non-occluded images in session 2 as the probe set. This database is more challenging since it contains 100 individuals compare to 10 in Yale B and 38 in Extended Yale B databases [25]. The precision and recall values for this data base are 99.38% and 99.29%, respectively.

7.7.1.3 Multi-PIE database

In the CMU Multi-PIE database, there are 750,000 images of 337 individuals in four different sessions with 15 pose and 19 illumination conditions and various facial expressions [38]. Some samples of the original and the preprocessed images of the first subject are shown in Fig 7.5c. In these experiments, all frontal pose images of 249 subjects with fourteen illumination conditions {0, 1, 3, 4, 6, 7, 8, 11, 13, 14, 16, 17, 18 and 19} and neutral expression in session 1 were used as the gallery set. Four probe sets include ten frontal images of even-number illumination conditions from sessions 2–4. The gallery and probe sets include 3486, 2030, 2300 and 2390 cropped images of size 96 × 72 pixels, respectively, in order to keep the consistency in comparison [36]. Table 7.5 shows recognition results on four probe sets. Precision and recall values are 97.59% and 97.54% for session 2, 96.13% and 96% for session 3, and 98.79% and 98.7% for session 4, respectively.

Table 7.5. Accuracy percentage of the face identification algorithm on the Multi-PIE database (%).

Method / Subset	SRC [25]	CRC [37]	LRC [29]	GRRC-l2 [36]	Proposed Method
Session 2	93.9	94.1	87.1	97.1	97.54
Session 3	90.0	89.3	81.9	96.8	96
Session 4	94.0	93.3	84.3	98.7	98.7

7.7.1.4 FRGC database

In order to evaluate our method on a highly challenging and large database, we use the images in the experiment 4 of the FRGC version 2 data set that contains images taken in uncontrolled illumination, with expression variation, out of focus blur and aging effect. FRGC 2.0.4 contains 12776 training images, 16028 target images and 8014 query images [39].We resized the images to 120×120 pixels. Some samples of the original and illumination insensitive images are shown in Fig 7.5d. In the first experiment, following the experimental settings in [5], one random image per person from the target set and query images were selected to make the gallery set of 466 images and the probe set of 466 images, respectively. The process was repeated 50 times, and the average recognition rate is shown in Table 7.6. The precision and recall values were calculated as 54.91% and 32.42%, respectively.

As a further challenging condition, in the second experiment on FRGC 2.0.4, we use one sample per individual in the target set as the gallery set of 466 images and all 8014 samples in the query set as the probe images. Table 7.7 illustrates the improved recognition results compared to the proposed method in [40]. The precision and recall values are 42.26% and 25.17%, respectively.

Table 7.6. Accuracy percentage of the first experiment of the face identification algorithm on the FRGC 2.0.4 database (%).

Method	Gabor-PCA [30]	LBP [26]	LOG-DCT [8]	OLHE [5]	Proposed Method
Recognition Accuracy (%)	13.73	11.99	15.82	20.53	32.62

Table 7.7. Accuracy percentage of the second experiment of the face identification algorithm on the FRGC 2.0.4 database (%).

Method	Gabor-PCA [30]	AGL [41]	NN-ADA-AM [40]	NN-ADA-RM [40]	Proposed Method
Recognition Accuracy (%)	9.57	15.49	18	18	23.3

7.7.2 Stability of the Algorithm versus Blur

The LPQ extracts the illumination and blur invariant features of image which makes the algorithm tolerant against lighting and blur effect. In order to evaluate the performance of our proposed method against blur, we impose different levels of out-of focus blur, approximated by a Gaussian kernel with 5 various standard deviations, $\sigma = 0, 2, 4, 8$ and 16 [42]. Figure 7.6 shows the blurred images of one sample in subset 4 of the Yale B database and their illumination normalized images. The first subset in this database is chosen as the gallery set and the remaining subsets are affected by blur and used as four probe sets. Figure 7.7 shows the recognition percentage versus the blur level for subsets 2-5. As shown, the break down point is at $\sigma = 8$, where the recognition rate falls down.

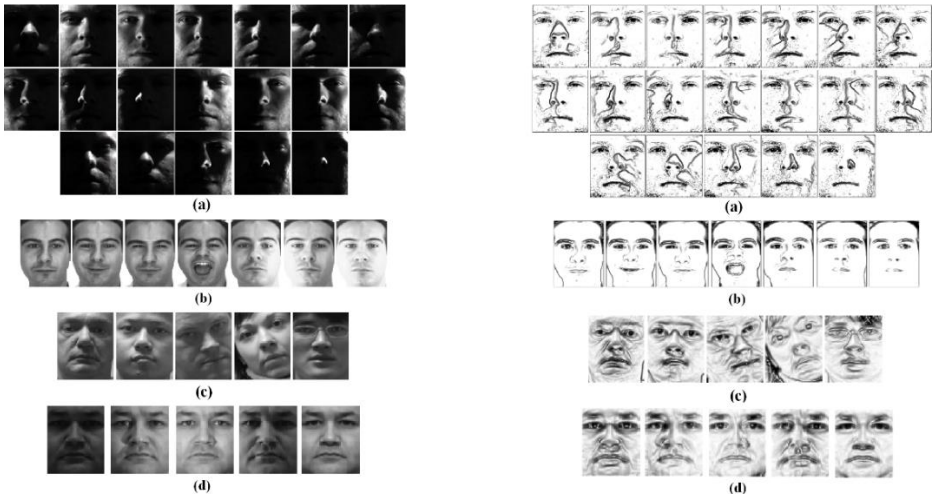

Figure 7.5. Sample images of different databases; left column: the original images, right column: the illumination invariant representation; rows from top to bottom: (a) Extended Yale B subset 5, (b) AR, (c) Multi-PIE and (d) FRGC 2.0.4.

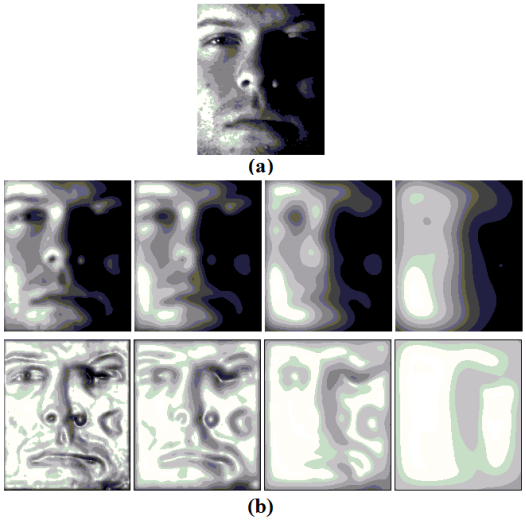

Figure 7.6. One example of the images in the Yale B, subset 4: (a) the original images, (b) first row; from left to right, the blurred images with Gaussian kernel of standard deviations, $\sigma = 2, 4, 8$ and 16, respectively, second row; their illumination normalized images.

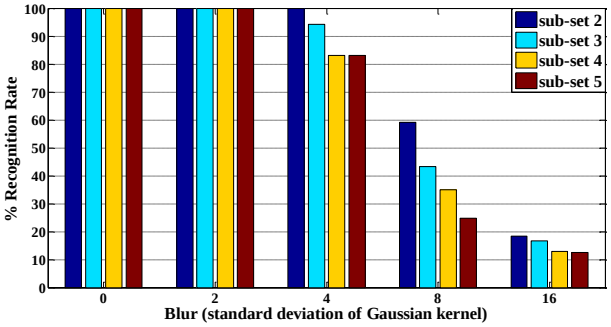

Figure 7.7. The recognition rate (%) of the algorithm for four subsets of the Yale B database versus different levels of the blur effect (standard deviation of Gaussian kernel).

7.8 CONCLUSION

Illumination variation affects the face appearance significantly and reduces the accuracy of the recognition system. In this chapter, a local illumination insensitive face identification algorithm is proposed, which is the combination of image preprocessing and two illumination insensitive local descriptors, LBP and LPQ. Illumination insensitive representation is composed based on the ratio

of gradient magnitude to the original image intensity value, which removes the luminance component superiorly. The preprocessed image is partitioned into small sub regions and LPQ and multi-resolution LBP, which are easy to implement and insensitive to monotonic variations, extract illumination insensitive local histograms. Score-level fusion combines local distances which are calculated by Chi-Square nearest neighbor classifiers to obtain the most probable match for the probe image. Decision-level fusion is employed to combine the results of two matching techniques. Entropy, class posterior probability and mutual information are utilized as the fusion weights. The recognition results on the Yale B, Extended Yale B, AR and very large and challenging Multi-PIE and FRGC databases show the superior behavior of the algorithm in severe illumination conditions, with a large number of subjects. The proposed method is computationally cost effective and easy to execute since it does not need training and image reconstruction.

7.9 REFERENCES

[1] P. H. Lee, S.W. Wu, Y. P. Hung, "Illumination compensation using oriented local histogram equalization and its application to face recognition," *IEEE Transactions on Image Processing*, vol. 21, (9) pp. 4280-4289, 2012.

[2] H. Han, S. Shan, X. Chen, W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, (6) pp. 1691-1699, 2013.

[3] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, "Face recognition under varying illumination using gradientfaces," *IEEE Transaction on Image Processing*, vol. 18, (11) pp. 2599-2606, 2009.

[4] B. Xu, T. P. Zhang, Z. W. Shang, "Multi-scale invariant abstracted under varying illumination," in *Proc. IEEE International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR12)*, Xian, 2012, pp. 28-32.

[5] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, "Subspace learning from image gradient orientations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, (12) pp. 2454-2466, 2012.

[6] B. Wang, W. Li, W. Yang, Q. Liao, "Illumination normalization based on Weber's law with application to face recognition," *IEEE Signal Processing Letters*, vol. 18, (8) pp. 462-465, 2011.

[7] A. Baradarani, Q. M. J. Wu, M. Ahmadi, "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering," *Pattern Recognition*, vol. 46, (1) pp. 57-72, 2013.

[8] W. Chen, M. J. Er, S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 36, (2) pp. 458-466, 2006.

[9] X. Tan, B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, (6) pp. 1635-1650, 2010.

[10] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, "Illumination invariant feature extraction and mutual information based local matching for face recognition under illumination variation and occlusion," *Pattern Recognition*, vol. 44, (10-11) pp. 2576-2587, 2011.

[11] T. Ahonen, A. Hadid, M. Pietikainen, "Face recognition with local binary patterns," *Springer Computer Vision*, vol. 3021, pp. 469-481, 2004.

[12] S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," in *International Conference on Pattern Recognition (ICPR12)*, Tsukuba, November 2012, pp. 1699-1702.

[13] A. Dhall, A. Asthana, R. Goecke, T. Gedeon, "Emotion recognition using PHOG and LPQ features," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG11)*, Santa Barbara, CA, 2011, pp. 878-883.

[14] D. Zang, "Illumination invariant object tracking based on multiscale phase," in *Proc. IEEE International Conference on Image Processing (ICIP10)*, Hong Kong, 2010, pp. 357-360.

[15] Y. Li, S. Shan, H. Zhang, S. Lao, X. Chen, "Fusing magnitude and phase features for robust face recognition," in *Proc. Asian Conference on Computer Vision (ACCV12)*, Daejeon, 2012, pp. 601-612.

[16] C. H. Chan, M. A. Tahir, J. Kittler, M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, (5) pp. 1164-1177, 2012.

[17] S. R. Zhou, J. P. Yin, J. M. Zhang, "LBP and LPQ based on Gabor filter for face representation," *Neurocomputing*, vol. 116, pp. 260-264, 2013,.

[18] V. Ojansivu, J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *Proc. International Conference on Image and Signal Processing (ICISP08)*, Cherbourg-Octeville, 2008, pp. 236-243.

[19] M. R. Banham, A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, (2) pp. 24-41, 1997.

[20] S. Jayaraman, S. Esakkirajan, T. Veerakumar, *Digital Image Processing*, New Delhi : Tata McGraw Hill, 2009.

[21] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, New Jersey : John Wiley and Sons, 2004.

[22] B. Topc, H. Erdogan, "Decision fusion for patch-based face recognition," in *Proc. IEEE International Conference on Pattern Recognition (ICPR10)*, Istanbul, Turkey, 2010, pp.1348-1351.

[23] S. Georghiades, P. N. Belhumeur, D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, (6) pp. 643-660, 2001. The YaleB face database, Available (online) at: <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>.

[24] K. C. Lee, J. Ho, D. Kriegman, "Acquiring linear sub spaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, (5) pp. 684-698, 2005. The Extended YaleB face database, Available (online) at:
<http://vision.ucsd.edu /~leekc/ExtYaleDatabase/ExtYaleB.html >.

[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, (2) pp. 210-227, 2009.

[26] T. Chen, X. S. Zhou, D. Comaniciu, T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, (9) pp. 1519-1524, 2006.

[27] M. Ezoji, K. Faez, "Use of matrix polar decomposition for illumination-tolerant face recognition in discrete cosine transform domain," *IET Image Processing*, vol. 5, (1) pp. 25-35, 2011.

[28] T. Zhang, B. Fang, Y. Yuan, et al., "Multiscale facial structure representation for face recognition under varying illumination," *Pattern Recognition*, vol. 42, (2) pp. 251-258, 1997.

[29] I. Naseem, R. Togneri, M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, (11) pp. 2106-2112, 2010.

[30] L. Chengjun, H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Transactions on Neural Networks*, vol. 14, (4) pp. 919-928, 2003.

[31] L. Chengjun, H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, (4) pp. 467-476, 2002.

[32] A. Martinez, R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998. Available (online) at: < http:// www2.ece.ohio-state.edu/~aliex/ARdatabase.html >.

[33] C. Y. Lu, H. Min, J. Gui, L. Zhu, Y. K. Lei, "Face recognition via weighted sparse representation," *Visual Communication and Image Representation*, vol. 24, (2) pp. 111-116, 2013.

[34] J. Qian, J. Yang, G. Gao, "Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction," *Pattern Recognition*, vol. 46, (10) pp. 2724-2739, 2013.

[35] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Transactions on Information Forensics and Security*, vol. 7, (3) pp. 954-965, 2012.

[36] M. Yang, L. Zhang, S. C. K. Shiu, D. Zhang, "Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary," *Pattern Recognition*, vol. 46, (7) pp. 1865-1878, 2013.

[37] L. Zhang, M. Yang, X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *Proc. IEEE International Conference on Computer Vision (ICCV11)*, Barcelona, 2011, pp. 471-478.

[38] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, (5) pp. 807-813, 2010.

[39] P. J. Phillips, P. Flynn, T. Scruggs, et al., "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, San Diego, 2005, pp. 947-954.

[40] M. Kana, S. Shana, Y. Suc, D. Xub, X. Chena, "Adaptive discriminant learning for face recognition," *Pattern Recognition*, vol. 46, (9) pp. 2497-2509, 2013.

[41] Y. Su, S. Shan, X. Chen, W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR10)*, pp. 3699-2706, 2010.

[42] P. Vageeswaran, K. Mitra, R. Chellappa, "Blur and illumination robust face recognition via Set-Theoretic characterization," *IEEE Transactions on Image Processing*, vol. 22, (4) pp. 1362-1372, 2013.

Chapter 8

Fusion of Global and Local Based Techniques for Face Recognition Algorithm under Degraded Conditions

Face recognition is a powerful biometric identification technique which is employed in diversified applications such as human computer interaction, passport control, video surveillance and law enforcement. The accuracies of face recognition techniques are degraded dramatically in real-world scenarios due to information corruption caused by occlusion, illumination variation, facial expression, head pose, aging effect, blur and low resolution. Occlusion and illumination variation cause crucial image degradation in face recognition which occur in applications such as video surveillance and uncontrolled imaging conditions where some parts the face area is covered by shadow or external obstacles such as scarf, helmet, sun glasses or another object [1]. Therefore, in some areas of the face image, distinctive facial information is corrupted which reduces the accuracy of recognizing individuals according to their face substantially. In order to tackle the illumination degradation, a variety of state of the art techniques have been employed in the literature. Preprocessing techniques employ the intensity level transformation or gradient based methods to remove the influence of lighting variation and enhance the face appearance [2, 3]. Histogram equalization (HE) and Gamma correction (GC) enhance the image appearance by

flattening the intensity distribution of image and modifying image gray values using a non-linear transformation, respectively [4, 5]. The proposed orientated local histogram equalization (OLHE) technique in [6], compensates illumination while encoding rich information on the edge orientations which is useful for face recognition. This technique performs exceptionally well when extreme lighting conditions occur. The proposed self-quotient image (SQI) in [7], is the ratio between the current image and the smoothed version of image. Whereas, in the reflection image model, the luminance component, $L(x, y)$, is the low frequency component of image and changes slowly, compared to the reflection component, $R(x, y)$, [2], In the proposed technique in [8] the ratio of directional image derivatives leads to the illumination insensitive representation. The authors modified their proposed approach into a multi-resolution analysis in [9]. They undertook the wavelet decomposition of directional derivatives in the illumination insensitive ratio. However, in the aforementioned gradient-based methods the image appearance is overexposed in severe illumination variation, thus in [10] the authors have recommended improvements by utilizing the tangent inverse of the ratio between magnitude of image gradient and the image intensity value, in order to remove the illumination artefacts. In [11] subspace learning techniques are applied on the gradient orientations. While the pixel intensity based subspace learning fails to estimate the low-dimensional data, the authors of [11] showed that using gradient orientation and a cosine-based correlation gives a more robust representation against illumination variation and occlusion. Weber's law proves that a noticeable change in a stimulus is a constant ratio of its original value [12]. In [13], the authors showed that the "Weber-face" (WF) which is captured using the arctangent of the ratio between the second order derivatives of image, using the Laplace operator, and current image intensity value, depends only on the reflectance component and thus leads to an illumination insensitive representation of face image. However, WF is in pixel-level and only the 8×8 neighbourhood around the centre pixel is taken into consideration. The authors improved their approach in [14] to a patch-level generalized form of WF where the mean intensity values of the outer and inner neighbors in a patch surrounding a pixel are taken into account. Also, the effects of inner and outer neighbors are weighted differently which is called weighted generalized WF (wGWF).

Wavelet transform is a viable illumination invariant preprocessing method. In [15] an image normalization technique is proposed based on the 1-level Daubechies10 wavelet transform by

decomposing image into frequency subbands. Contrast and edge enhancement is applied on the approximation and detailed coefficients, respectively, which are fed into the wavelet reconstruction. In the wavelet based illumination invariant image processing method proposed in [16], discrete wavelet transform (DWT) is applied on the image logarithm and the approximation coefficients which contains low frequency information are set to zero and followed by reconstruction and exponentiation. However, DWT is not shift invariant, which means the coefficients are changed by image translation, and suffers from poor directional selectivity for diagonal features [17]. In [18] a double-density dual-tree complex wavelet transform (DD-DTCWT), which is shift invariant, complex-valued, directional selective and has perfect reconstruction is applied to the preprocessed image, using the normalized logarithm function. The high frequency subbands are thresholded and reconstructed image is passed through the maximum filter and edges amplification in order to extract the refection component. In the proposed method in [19], an adequate number of low frequency coefficients of discrete cosine transform (DCT) are eliminated, by setting them to zero, in order to reduce the illumination effect. Illumination insensitive feature extractors such as local binary pattern (LBP) [20], local phase quantization (LPQ) [21] and Gabor filter [22, 23] which derive the characteristics of image, insensitive to lighting condition, can also be utilized in order to cope with the illumination variation. Nevertheless, compound approaches lead to superior performance. In the method proposed in [24] a preprocessing chain, which is the fusion of GC, difference of Gaussian (DoG) filter, and HE, is carried out. Moreover, local ternary pattern (LTP), the extension of LBP, as an illumination insensitive feature extractor derives the local texture characteristics of face image.

Partial occlusion on the face area which is caused by an external object such as sun glasses, scarf, hand or another subject's face [1], corrupts some information useful in identification of the individual and reduces the accuracy of face recognition algorithm. Holistic-based subspace learning approaches which extract features from the whole facial area, are not appropriate techniques to cope with occlusion problem [25]. The occluded portion of face affects the entire extracted characteristics which leads to misclassification of the unknown identities. However, by utilizing local-based descriptors and reducing the influence of non-facial information related to the occluded parts of the face on the classification, the partial occlusion challenge would be tackled substantially. Therefore, the occluded parts should be detected and lower weights be

devoted to their extracted features in combining the local descriptions [25, 26]. Local-based techniques are also useful in order to exclude the effect of artefacts due to illumination variation or facial expression. In [27] a local-based probabilistic subspace approach was employed on the image sub blocks and each image sub region in the training set is projected into its own local eigenspace. The class label attributed to the probe image is the maximum argument of the summation of local probabilities. In the technique proposed in [28], nearest neighbor classifier detects the occluded image sub blocks first. Then, selective local non-negative matrix factorization (LNMF) approach is employed and for each LNMF basis, the portion of its energy which contributes to occluded regions is calculated. If it is less than the threshold, the considered LNMF basis is assumed as a non-occluded basis. The training and testing sets are projected to the new subspace based on the occlusion-free bases. The authors of [29, 30] applied pseudo Zernike moment (PZM) and Gabor filter, respectively, as local descriptors on the image sub blocks. The local features are concatenated and their contribution in the feature combination is calculated using the adaptive occlusion map and local entropy as local weights of image sub regions. DCT is applied on image sub blocks in [31] and a small number of frequency coefficients are selected in a Zigzag manner. Local features are combined and counterbalanced using the standard deviation and an alignment method is applied on the face images at classification stage to deal with the occlusion problem. A sparse representation is applied in [32], where the test sample is a linear combination of training samples of the same individual plus the error caused by occlusion with small number of non-zero elements. Partial occlusion error is assumed to have a sparse nature since it corrupts a fraction of pixels. The l_1-norm solution is sparse enough when the dimension of features is sufficiently large [32]. However, since the occluded images are not included in the training set, reconstruction error of the occluded regions in the probe image is big. Therefore, in order to tackle the occlusion effect, in [33] a representation residual map is obtained using the training samples and calculating the distribution of coding residuals. The outliers in the probe set are detected using an adaptive threshold on the calculated map and excluded from the recognition process where the probe image is coded over a dictionary. A block-based face recognition algorithm is proposed in [34] with one gallery sample per subject. Uniform LBP extracts local histograms of equal-sized and non-overlapping sub blocks of image. Chi-square based nearest neighbor classifiers find the local class labels of the probe image. The authors of [34] utilized

weighted majority voting scheme as a decision fusion strategy and local entropy is applied to lower the effect of partially occluded sub regions. Local Gabor binary pattern histogram (LGBPH) technique is proposed in [35] by concatenating the LBP histograms of the magnitude of Gabor features in image sub blocks. Kullback–Leibler divergence (KLD)-is employed to calculate the probability of occlusion in the image sub blocks by comparing the local KLD's with a threshold, which is used to weigh the local histograms in the feature concatenation [35]. However, determining the threshold requires a training stage and its value affects the recognition accuracy. Nevertheless, regarding the psychological point of view about the human perception of people identities based on their face image, both global and local information is taken into consideration [36, 37]. Therefore, fusion of global-based and local-based approaches, leads to improved recognition accuracy. In [38] a multi sub-region based correlation filter bank technique called MS-CFB was proposed where the extraction and combination of features of image sub regions were unified and the local correlations are used to extract an optimized global correlation for each image. The intensity values of the sub regions pixels are concatenated and class specific correlation filter banks are calculated for the training samples of different classes in the training set to extract the features of training and test images. Directionality and texture characteristics of the face image at the global and local levels are combined in [39]. Local polynomial approximation (LPA) filter is applied on the face images and scale optimization at each direction is carried out to obtain directional images at different scales and directions. Modified LBP is applied on each directional image and the LBP images are partitioned into sub regions at four partitioning levels. The local histograms of image partitions of four levels are concatenated. The global feature vector of image is the combination of histograms of directional images. Finally LDA is utilized as the dimensionality reduction strategy.

In the proposed face recognition algorithm in this chapter, a hierarchical structure of global and local recognition strategies is applied, based on the human perception in recognition of unknown face. Figure 8.1 shows the block-diagram of the proposed algorithm. The combination of multi-scale Weber's technique and wavelet based preprocessing is applied on the face image in order to find the illumination insensitive representation of image. Gabor filter bank is applied on the whole face image to extract features which are insensitive against intra-subject variations, directionally and spatially selective and space-frequency localized. Furthermore, through the local

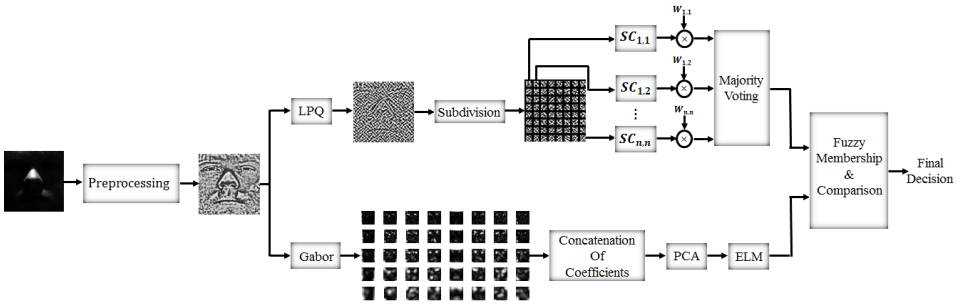

Figure 8.1. Block diagram of the proposed face recognition algorithm.

based approach, the illumination and blur invariant texture descriptor, called local phase quantization (LPQ) [21], is applied on the preprocessed image and the resultant face image is partitioned into sub blocks. Local histograms of different sub blocks are fed into the local classifiers, based on sparse representation technique [32], and combined using a majority vote decision fusion. However, in order to reduce the effect of corrupted sub blocks with less discrimination power, we employ the multiplication of the sparse classifier residual and Kullback–Leibler distance [35] as the local weights and differentiate the blocks votes in concluding a decision on the identity of the face image. Finally, final class label of the unknown individual is obtained by comparing fuzzy membership (FM) of the block-based and global-based results and finding the class with maximum membership value.

The rest of chapter is organized as follows. In section 8.1 the illumination insensitive preprocessing technique is explained mathematically. Section 8.2 describes the proposed face recognition algorithm in details. Simulation parameters, face databases and implementation results are shown in section 8.3. This chapter is concluded in section 8.4.

8.1 ILLUMINATION INVARIANT PREPROCESSING

In the proposed method of image normalization to deal with the degradation due to illumination variation, we proposed a combination of two strategies as follows.

8.1.1 Multi-Scale Weber

The proposed method is the multi-scale extension of Weber's technique [13]. Based on the psychological law proposed by Ernst Weber in 1834, a noticeable variation in a stimulus is a

constant ratio of its original value [12]. In the proposed method in [13], the illumination insensitive representation of the face image, Weber-face, is calculated by the tangent inverse of the ratio between a noticeable difference of each image pixel and its 8 close neighbours and the current pixel intensity value, which is the second order derivative of image, as follows [13],

$$WF = tan^{-1}\left[\alpha \sum_{a\in A}\sum_{b\in A}\frac{i(x,y)-i(x-a\Delta x,y-b\Delta y)}{i(x,y)}\right], \tag{8.1}$$

where $i(x,y)$is the image gray value, α is the parameter to adjust the difference between pixels and A=[-1, 0, 1]. According to the reflection model of the image, we have the multiplication of two components,

$$i(x,y) = r(x,y) \times l(x,y), \tag{8.2}$$

where $r(x,y)$ and $l(x,y)$ are the reflection and luminance component, respectively. Since luminance is the low frequency component of image with slow variation, we have the following approximation,

$$l(x - a\Delta x, y - b\Delta y) \cong l(x,y). \tag{8.3}$$

Therefore, the Weber-face in (8.1) is not dependant on the illumination variation and it varies regarding the reflection component,

$$\begin{aligned} WF &= tan^{-1}\left[\alpha \sum_{a\in A}\sum_{b\in A}\frac{l(x,y).[r(x,y)-r(x-a\Delta x,y-b\Delta y)]}{l(x,y).r(x,y)}\right] \\ &= tan^{-1}\left[\alpha \sum_{a\in A}\sum_{b\in A}\frac{[r(x,y)-r(x-a\Delta x,y-b\Delta y)]}{r(x,y)}\right]. \end{aligned} \tag{8.4}$$

Tangent inverse is utilized to prevent the WF ratio from becoming too large and thus limits the output noise [13]. In order to reduce the effect of noise on the WF and reduce the artifacts of shadow boundaries, the image is first smoothened by a Gaussian filter with standard deviation σ, 0.8 in this chapter [13]. In the proposed technique in this chapter, we employed the multi-scale Weber analysis in order to enhance the normalization method to a scale invariant representation.

The differences between each pixel and its neighbours are utilized with different neighborhood sizes and the results are added to each other. In this chapter we applied three neighbourhood sizes of 3×3, 5×5 and 7×7 pixels and 3 as α value. The values of standard deviation and α parameter are obtained according to the exhaustive search results.

8.1.2 Image Normalization Based on DD-DT-CWT

Discrete wavelet transform is a powerful function in order to locally analyze image in multiresolution manner in time and frequency domain simultaneously and separate the image frequency subbands. Image can be displayed as of different scaling and wavelet basis functions [16]. A one level DWT extracts the approximation (produce the low-low subband) and detail information (generate the low-high, high-low and high-high subbands) of image. Luminance component is the low frequency component of image which lies in the LL subband. Whereas, high frequency subbands include the horizontal, vertical and diagonal details of image such as edge and structure information [16]. Therefore, if the contrast of LL subband is enhanced, the illumination variation effect would be reduced significantly in the reconstructed image using inverse DWT. However, the type of wavelet transform is a critical factor in the yielded normalized image. DWT suffers dramatically from the lack of shift (image translation) invariance, directional selectivity and perfect reconstruction. The combination of double-density DWT, which is a shift invariant and directionally selective wavelet transform, and dual-tree DWT, which is a complex-valued wavelet useful for image modeling and denoising, employs the advantages of both DWTs in the form of double density dual tree discrete wavelet transform (DD-DT-DWT) [40]. A 2-dimensional DD-DDT-DWT is constructed of two scaling functions, $\phi_h(t)$ and $\phi_g(t)$, and four wavelet bases, $\psi_{h,\mathrm{i}}(t)$ and $\psi_{g,\mathrm{i}}(t)$ where $\mathrm{i} = 1$ and 2. Thus, two oversampled 2-dimensional double density DWT filter banks in parallel are applied for imaginary and real parts as shown in Fig 2 [40, 18]. Two wavelets are offset by one half and form a Hilbert transform, $\mathcal{H}[.]$, pair as follows,

$$\begin{cases} \psi_{h,1}(t) = \psi_{h,2}(t - 0.5). \\ \psi_{g,1}(t) = \psi_{g,2}(t - 0.5). \\ \psi_{g,1}(t) = \mathcal{H}[\psi_{h,1}(t)]. \\ \psi_{g,2}(t) = \mathcal{H}\big[\psi_{h,2}(t)\big]. \end{cases} \tag{8.5}$$

The analysis scaling and wavelet functions are as follows,

$$\begin{cases} \phi_h(t) = \sqrt{2}\ \sum_m h_0(m)\, \phi_h(2t-m). \\ \psi_{h,1}(t) = \sqrt{2}\ \sum_m h_1(m)\, \psi_h(2t-m). \\ \psi_{h,2}(t) = \sqrt{2}\ \sum_m h_2(m)\, \psi_h(2t-m). \end{cases} \quad (8.6)$$

The analysis scaling and wavelet set of $\phi_g(t)$, $\psi_{g,1}(t)$ and $\psi_{g,2}(t)$ are defined as above [40]. The synthesis filters are time reverse of analysis filters. In order to hold the Hilbert pair constraint, $\psi_{g,1}(t) = \mathcal{H}[\psi_{h,1}(t)]$ and $\psi_{g,2}(t) = \mathcal{H}[\psi_{h,2}(t)]$. Their Fourier transform are shown as follows,

$$\Psi_{g,1}(\omega) = \begin{cases} -j\left[\Psi_{h,1}(\omega)\right] & for \quad \omega > 0 \\ +j\left[\Psi_{h,1}(\omega)\right] & for \quad \omega < 0 \end{cases}, \quad \Psi_{g,2}(\omega) = \begin{cases} -j\left[\Psi_{h,2}(\omega)\right] & for \quad \omega > 0 \\ +j\left[\Psi_{h,2}(\omega)\right] & for \quad \omega < 0 \end{cases}, \quad (7.7)$$

$\Psi_{g,i}(\omega)$ and $\Psi_{h,i}(\omega)$ are the Fourier transforms of our wavelet basis functions where $i = 1,2$. Similarly, the Hilbert pair conditions are defined for the synthesis filters. Decomposition of image using DD-DT-DTW yields more frequency subbands compare to DWT (2 approximation or scaling subbands and sixteen highpass or detail subbands for each level) as shown in Fig. 8.2 and also extracts image features in several local orientations [40].

Similarly, a 2-dimensional double density dual tree complex wavelet transform (DD-DDT-CWT) is constructed of four oversampling 2-dimensional double density DWT filter banks in parallel which yields 4 scaling and 32 detail subbands in each decomposition level (more specific detail information compare to DD-DDT-DWT). The 36 decomposition samples of the low and high frequency subbands are shown in Fig 8.3. In the proposed method, we have applied DD-DT-CWT on the normalized image from the previous section. In order to further reduce the effect of illumination variation, and enhance the structure details of image, similar to the technique proposed in [15], we use the HE approach to enhance the contrast of low frequency subbands and enlarge the values of high frequency coefficients by a scalar multiplication which is greater than 1 and is determined regarding the average of recognition accuracy through an exhaustive search, equal to 2.5. Image reconstruction is carried out using the synthesis filters. Figure 8.4 shows

some sample images with severe illumination variation in Extended Yale B database and their illumination insensitive representation using the proposed technique.

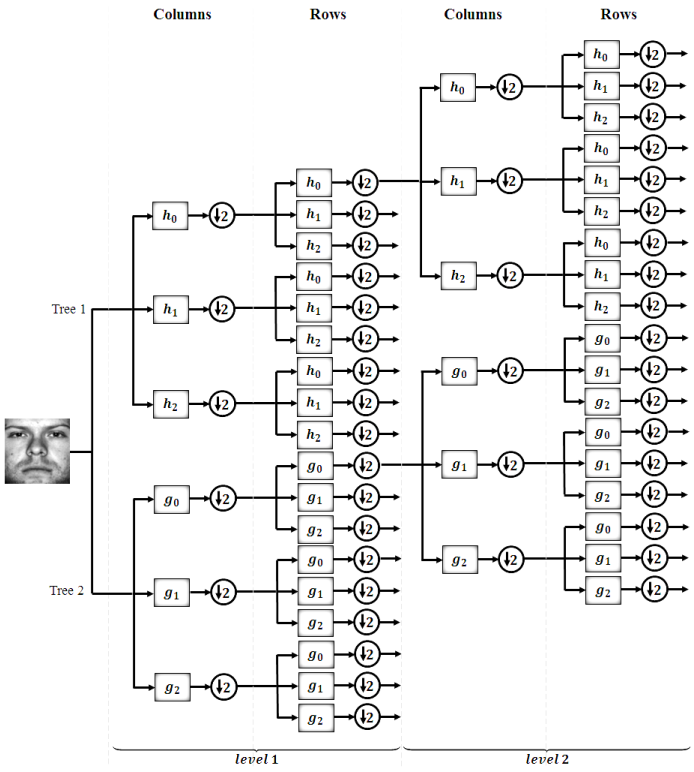

Figure 8.2. Decomposition structure of 2-dimensional DD-DT-DWT.

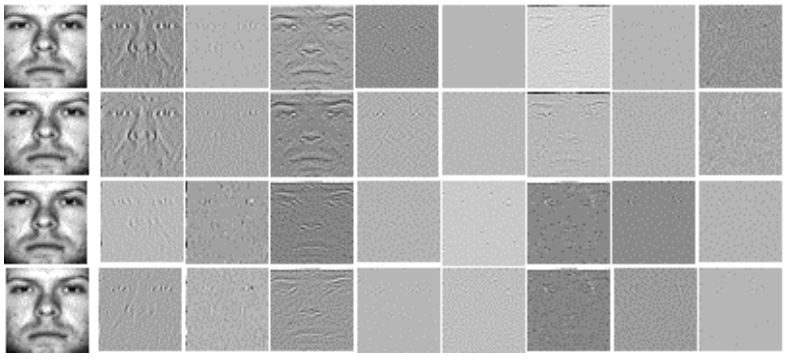

Figure 8.3. The frequency subbands of applying DD-DT-CWT on the face image: (a) low frequency (scaling) subbands and (b) high frequency (detail) subbands.

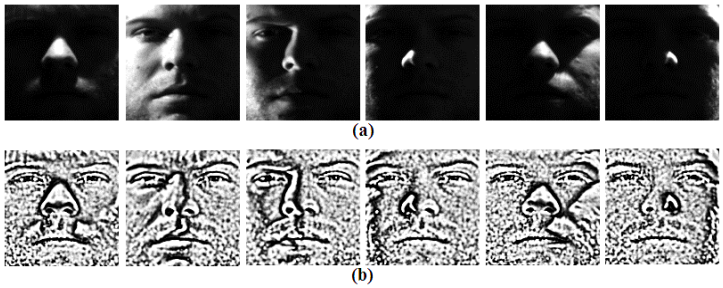

Figure 8.4. (a) Sample images with severe illumination variation from Extended Yale B database and (b) illumination insensitive representations using the proposed technique.

8.2 PROPOSED FACE RECOGNITION ALGORITHM

According to the psychological studies on human perception in identifying individuals by looking at their face, both global and local information is taken into consideration [36, 37]. Therefore, in the proposed face recognition algorithm, as shown in Fig. 8.1, we use a hierarchical structure of global and local face recognition algorithms by fusion of the decisions on the class label of the probe image, which are made by a global-based identification technique, and a block-based method, on the illumination normalized image. The 2D-Gabor filter is applied on the whole face image in order to extract face characteristics which are illumination invariant and localized in space-frequency domain and directionally and spatially selective. Gabor technique outperforms holistic feature extractors in order to tackle the local image degradation. Moreover, for block-based technique, we apply the local phase quantization (LPQ) approach which is illumination and blur invariant texture descriptor to extract feature vectors of image sub blocks separately. We do not apply Gabor in the block-based technique since LPQ improves the recognition results significantly compared to Gabor, regarding the result of experimental exhaustive search. However, it is not as effective as Gabor in global face recognition due to the lower number of extracted features out of the whole face. Moreover, these two feature extraction techniques provide complementary information of shape and texture of the face image. Similarly, we apply different classifiers for the global and local based recognition, extreme learning machine (ELM) and sparse-based classifier (SC), respectively.

8.2.1 Global-Based Approach

8.2.1.1 Gabor feature extraction

2D-Gabor wavelets are desirable feature extraction techniques in pattern recognition applications due to the similarity between the Gabor kernels and the response of 2D receptive field profiles of the mammalian cortical simple cell. They are able to extract image salient characteristics with spatial locality, orientation and spatial frequency selectivity and robustness against local deformations due to illumination variation, facial expression [41-43]. Gabor filter is constructed by the modulation of a sinusoid function using a Gaussian kernel [22, 23] as follows,

$$\psi_{s,o}(x,y) = \frac{{f_s}^2}{\pi\delta\theta} \times e^{-\left[\left(\frac{{f_s}^2\hat{x}^2}{{\delta_x}^2}\right)+\left(\frac{{f_s}^2\hat{y}^2}{{\delta_y}^2}\right)\right]} \times e^{-(j2\pi.f_s.\hat{x})}, \tag{8.8}$$

$$\begin{cases} s = 0,1,2,\dots,S_{max}-1 \\ o = 0,1,2,\dots,O_{max}-1 \\ \hat{x} = +x\cos(\beta_o) + y\sin(\beta_o) \\ \hat{y} = -x\sin(\beta_o) + y\cos(\beta_o) \end{cases}, \tag{8.9}$$

$\beta_o = (o\pi)/8$, $f_s = f_{max}/\left(\sqrt{2}\right)^s$ and f_{max} is the maximum frequency of the filter bank where γ and μ are the sharpness of x and y axis, respectively, and equal to $\sqrt{2}$. S_{max} and O_{max} are the maximum number of scale and orientation, respectively. Feature extraction is carried out by convolution of the face image and the Gabor filter bank, which concludes to a set of $S_{max} \times O_{max}$ coefficients per image pixel. All Gabor coefficients are concatenated in the form of a large vector. To reduce the dimension of feature vector and computational complexity, principal component analysis (PCA) as a dimensionality reduction subspace technique is applied [44]. We just consider the magnitude of Gabor coefficients since Gabor phase is not shift invariant and changes linearly with small translation [41].

8.2.1.2 ELM classification

Feed-forward neural networks are applied extensively as a classification strategy in pattern recognition applications due to their capability in approximating the nonlinear mappings in the

data [45]. In order to tune the weights and biases of the network, traditional learning mechanisms such as gradient decent method are employed. However, due to very slow iterative tuning by gradient decent technique and its convergence into local minima, feed-forward neural network suffers from slow learning and poor scalability. Extreme learning machine, as a learning algorithm for the single hidden layer feed-forward neural network, is an extremely faster technique. ELM classifier is the generalized single hidden layer neural network with random hidden nodes and determined hidden layer weights without iterative weight tuning [45]. For L distinct training samples, the single hidden layer feed-forward neural network with N_h random hidden neurons, M input and K output nodes is modeled as follows [45],

$$\sum_{j=1}^{N_h} \bar{\gamma}_j F_j(\bar{x}_i) = \sum_{j=1}^{N_h} \gamma_j F\left(\bar{w}_j . \bar{x}_i + b_j\right) = \bar{o}_i , \quad i = 1, 2, \dots, L, \tag{8.10}$$

where $\bar{x}_i = [x_{i1}, x_{i2}, \dots, x_{iM}]^T$ and $\bar{y}_i = [y_{i1}, y_{i2}, \dots, y_{iK}]^T$ are the input and output nodes and $F(.)$ is the activation function of the network. b_j is the threshold of j^{th} hidden node and $\bar{w}_j = \left[w_{j1}, w_{j2}, \dots, w_{jM}\right]^T$ and $\bar{\gamma}_j = \left[\gamma_{j1}, \gamma_{j2}, \dots, \gamma_{jK}\right]^T$ denote the weight vectors between the j^{th} hidden node and the input and output nodes, respectively. L samples can be approximated to have zero error means such that,

$$\sum_{j=1}^{N_h} \gamma_j F\left(\bar{w}_j . \bar{x}_i + b_j\right) = \bar{y}_i, \tag{8.11}$$

where (8.11) can be denoted as follows,

$$H\Upsilon = Y, \tag{8.12}$$

where $\Upsilon = \left[\bar{\gamma}_1{}^T, \bar{\gamma}_2{}^T, \dots, \bar{\gamma}_{N_h}{}^T\right]^T$, $\mathrm{Y} = \left[\bar{y}_1{}^T, \bar{y}_2{}^T, \dots, \bar{y}_L{}^T\right]^T$ and H is the hidden layer matrix, the l^{th} column of which is the output of l^{th} hidden node. It is proven in [45] that if $F(.)$ is differentiable infinitely, we can assign random values to the weights and biases and the required hidden layer nodes is $N_h \leq L$. Therefore, in the ELM technique, rather than tuning the weights and biases iteratively in gradient descent method, they are randomly assigned in the beginning of learning. Then, H is calculated and output weights are obtained through the following minimum norm least

squares solution of (8.12),

$$\widehat{\Upsilon} = H^{\dagger}Y, \tag{8.13}$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse of H [45].

8.2.2 Block-Based Approach

8.2.2.1 Illumination insensitive LPQ feature extraction

Local phase quantization (LPQ) is a popular histogram-based local texture descriptor in the computer vision and pattern recognition area. This technique is based on the analysis of phase information in a local image window at selected frequencies and phase quantization concludes in an efficient detailed illumination insensitive face representation [21, 46, 47]. Moreover, LPQ is blur insensitive. Blur effect is a degradation due to object motion or out of focus of camera in uncontrolled imaging conditions. The spatial blur model in the frequency domain is as follows,

$$B(u,v) = I(u,v).P(u,v), \tag{8.14}$$

where $B(u,v)$, $I(u,v)$ and $P(u,v)$ are discrete Fourier transform (DFT) of blurred image, $b(x,y)$, original image intensity, $i(x,y)$, and point spread function (PSF) of blur, $p(x,y)$, respectively. For a local neighbourhood around each image pixel, $W(x,y)$ at pixel position (x,y), we calculate the short time Fourier transform (STFT),

$$B(\alpha,\beta,u,v) = \sum_{y\in W}\sum_{x\in W} b(x,y).W^{*}(x-\alpha,y-\beta).e^{-j2\pi(ux+vy)}, \tag{8.15}$$

where α and β are the spatial position and u and v are the spatial frequency of the window, respectively [48]. The PSF of the motion, atmospheric turbulence or out of focus blur, can be assumed as centrally symmetric functions at small frequencies bellow the first zero crossing. Therefore, their frequency spectrum is a Sinc or Gaussian function and positive at those frequencies [49, 50]. Therefore, STFT of PSF, $P(u,v)$, at low frequencies within the bandwidth of PSF, is positive with zero phase. Thus, local analysis of STFT phase at small frequencies,

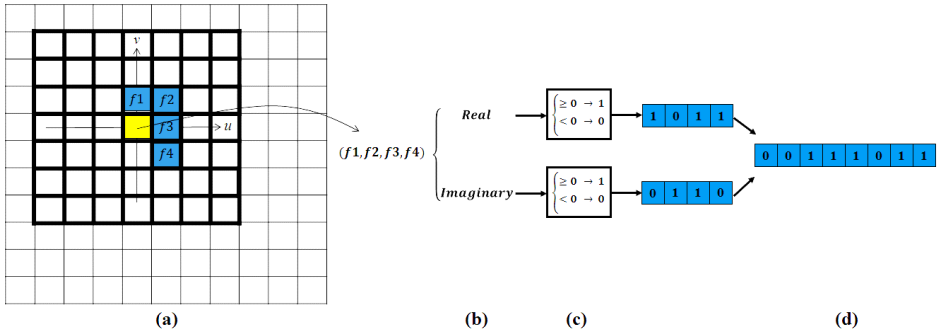

Figure 8.5. LPQ technique on each image pixel: (a) STFT coefficients at four specific frequencies inside a window at the pixel position, (b) finding the real and imaginary of the four frequency coefficients, (c) binary quantization based on the sign of real and imaginary values and (d) constructing an 8-bit stream for the corresponding pixel.

$[u_1, v_1] = [0, a], [u_2, v_2] = [a, a], [u_3, v_3] = [a, 0]$ and $[u_4, v_4] = [a, -a]$, is blur insensitive as follows [49],

$$< B(u, v) =< I(u, v), \tag{8.16}$$

$a = 1/r$, where $r \times r$ pixels is the size of local neighbourhood, $r = 7$ in this work. The following binary quantization of the sign of real and imaginary parts of four frequency coefficients leads to an illumination and blur insensitive description,

$$q_t = \begin{cases} 1 & if\ c_t \geq 0 \\ 0 & if\ c_t < 0 \end{cases}, \tag{8.17}$$

where c_t is the t^{th} element of the 8-component vector of real and imaginary parts of four frequency coefficients at each image pixel. In this work, in order to increase the tolerance against the mean level of illumination and illumination contrast, the real and imaginary transformed images are normalized to zero mean [46]. Figure 8.5 shows the LPQ approach for an image pixel.

8.2.2.2 Image partitioning

In the proposed method, followed by LPQ technique and obtaining the decimal values of image pixels corresponding to their 8-bit streams, we have divided the face image, i, of size $M \times M$ pixels, into sub blocks of appropriate size. According to the results which are obtained through an exhaustive search, we choose non-overlapped and equal-sized square blocks of size $w \times w$ pixels. Thus, the total number of sub blocks in an image is $N_b = \frac{M}{w} \times \frac{M}{w}$, in this work M=128 and $w = 8$. The sub block at position r and s is denoted as follows [29],

$$b_{rs}(u,v) = i(w(r-1)+u, w(s-1)+v), \quad u, = 1,2,\dots,w \quad and \quad r,s = 1,2,\dots,\frac{M}{w}. \tag{8.18}$$

By using the LPQ decimal value of image pixels, a histogram of 2^8 bins is assigned to each image sub-block. The classification technique is applied on the sub regions separately and the local class labels are combined by utilizing a decision-fusion strategy.

8.2.2.3 Sparse classification (SC)

The constructed LPQ histograms for image sub regions are fed into local classifiers based on the sparse representation approach where the samples belong to the same class of subject lie on a linear subspace [32].

The b^{rs} sub block of a gallery sample is defined as a vector of extracted LPQ features as follows,

$$g_{j,n}^{b_{rs}} = \begin{bmatrix} f_1^{g_{j,n}^{b_{rs}}} \\ f_2^{g_{j,n}^{b_{rs}}} \\ \vdots \\ f_m^{g_{j,n}^{b_{rs}}} \end{bmatrix}, \qquad n = 1,2,\dots,N_j, \tag{8.19}$$

where $f_m^{g_{j,n}^{b_{rs}}}$ is the m^{th} histogram bin of the b_{rs} block of the n^{th}gallery sample contained in the j^{th} class with N_j samples. Thus, we denote the set of gallery samples for block b_{rs} from j^{th} class as follows,

$$G_j^{b_{rs}} = \left[g_{j,1}^{b_{rs}}, g_{j,2}^{b_{rs}}, \dots, g_{j,N_j}^{b_{rs}}\right]. \tag{8.20}$$

The whole gallery sample dictionary for block b_{rs}, is obtained by concatenating the gallery sets of all classes of individuals,

$$G^{b_{rs}} = \left[G_1^{b_{rs}}, G_2^{b_{rs}}, \dots, g_{N_c}^{b_{rs}}\right]. \tag{8.21}$$

The b^{rs} sub block of a probe sample, p, which belongs to the k^{th} class, is represented as a linear combination of the b^{rs} sub blocks of all gallery samples corresponding to the same class,

$$p_k^{b_{rs}} = G^{b_{rs}}.\mathrm{B}^{b_{rs}} = \sum_{l=1}^{N_k} g_{k,l}^{b_{rs}}.\beta_{k,l}^{b_{rs}}, \tag{8.22}$$

where $\mathrm{B}^{b_{rs}}$ is the coefficient vector for block b_{rs} of the probe image whose entries are zero except those which are corresponding to the k^{th} class,

$$B^{b_{rs}} = \left[0, 0, \dots, 0, \beta_{k,1}^{b_{rs}}, \beta_{k,2}^{b_{rs}}, \dots, \beta_{k,N_k}^{b_{rs}}, 0, 0, \dots, 0\right]. \tag{8.23}$$

We can also include the effect of noise by modifying (8.22) and adding the noise term τ. The energy of noise is limited as $\|\tau\|_2 < \varepsilon$ [32],

$$p_k^{b_{rs}} = G^{b_{rs}}.\mathrm{B}^{b_{rs}} + \tau. \tag{8.24}$$

The coefficient vector encodes the class label of the sub-block of the probe image by solving the following l_1-minimization problem [51],

$$(l_1): \quad \hat{B}^{b_{rs}} = argmin\left\|B^{b_{rs}}\right\|_1 \qquad while \ \left\|G^{b_{rs}}.\mathrm{B}^{b_{rs}} - p_k^{b_{rs}}\right\|_2 \leq \varepsilon. \tag{8.25}$$

Due to the effect of noise, there might exist some non-zero coefficients in the $B^{b_{rs}}$ vector which are related to other classes of subject. In order to reduce the misclassification probability, $\delta_k(.)$ is utilized to select just the coefficients associated with k^{th} class, thus its non-zero elements are just corresponding to the k^{th} class. Therefore, the sub-block of the probe image can approximately be reconstructed as $\hat{p}_k^{b_{rs}} = G^{b_{rs}}.\delta_k(\hat{B}^{b_{rs}})$. The local decision on the class label of probe image sub-region, $d(r,s)$, is calculated by finding the argument which minimizes the residual between $p_k^{b_{rs}}$ and its approximation as follows [32],

$$d(r,s) = argmin_k\{r_k(p_k^{b_{rs}})\}, \tag{8.26}$$

where

$$r_k(p_k^{b_{rs}}) = \|p_k^{b_{rs}} - G^{b_{rs}}.\delta_k(\hat{B}^{b_{rs}})\|_2. \tag{8.27}$$

The number of gallery samples per subject affects the performance of sparse classifier. It should be sufficient to represent the probe sample under degraded conditions sparsely enough [32]. However, in real world applications a small number of samples per individual is available. Moreover, decreasing the number of gallery samples, reduces the computational cost of classification. In this chapter, we evaluated the effect of gallery size on the performance of the algorithm for 532 images of 38 subjects in subset 4 of Extended Yale B and 700 non-occluded images of 38 individuals in session 2 of AR databases (description of the databases in the following experimental results in section 4). By reducing the number of gallery samples per subject from 7 to 1, Fig 6 shows its effect on the recognition accuracy of the proposed technique. As shown in Fig 6, the number of gallery samples is more critical for AR database

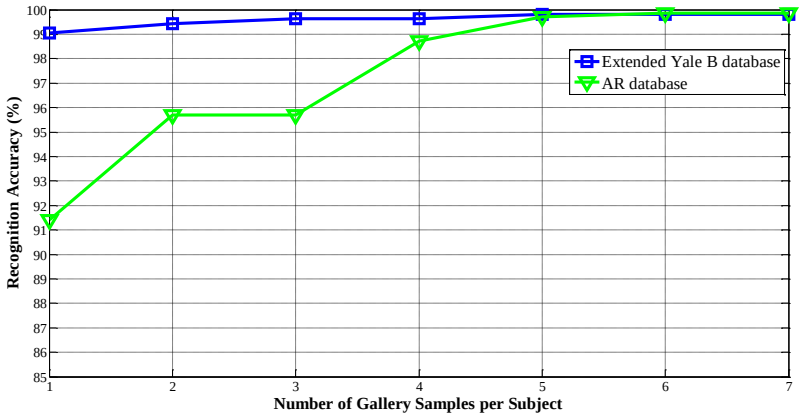

Figure 8.6. Recognition accuracy (%) of the proposed face recognition technique versus number of gallery samples per individual: a) subset 4 of the Extended Yale B database and b) non-occluded images of session 2 in AR database.

where the number of subjects is larger and also the probe samples are affected by more degraded conditions.

8.2.2.4 Local weighting

In order to further reduce the effect of corrupted sub blocks either by partial occlusion or shadow artifacts caused by illumination variation, we proposed a weighting scheme to devote weights to the image partitions regarding their discrimination power in identification of the subject. Therefore, the effects of corrupted regions are reduced at decision fusion in the following section on the final class label. The proposed local weight is the following multiplication of the normal values of two different weights, sparse residual weight (SRW) and Kullback-Lebler weight (KLW).

$$W(r,s) = SRW(r,s) \times KLW(r,s). \tag{8.28}$$

In the proposed SC technique in section 3.2.3 the local class label of the probe image sub block is obtained by finding the argument which minimizes the residual value in (8.27). In this work we use the residual value of the attributed class to the probe sub block to calculate the local weight

belonging to that sub block. The sub block which is not highly affected by degraded conditions, should have more influence on the final recognition of the probe face by appropriating higher value of local weight. Therefore, the local SRW is calculated as follows.

$$SRW(r,s) = 1 - r_k\left(p_k^{b_{rs}}\right). \tag{8.29}$$

Also, we calculate the Kullback–Leibler (KL) distance of the image sub blocks using their LPQ histograms. We take the actual values of local distances as the local weights and skip the probability computation and thresholding stages in [35] since selecting the value of threshold can significantly affect the identification accuracy and requires a training step in the recognition algorithm. Therefore, it would not be integrable to the new databases and could not be used in real-time applications and needs a large number of training samples which are not available in real world applications.

The KL distance between two densities $A(x)$ and $B(x)$ is as follows,

$$KL\_distance\{A(x), B(x)\} = \int_{-\infty}^{+\infty} A(x)\frac{A(x)}{B(x)}dx. \tag{8.30}$$

In our proposed method, we have calculated the class-average image for each individual class, which is the average image of all gallery samples belonging to the same class, and extracted the local histograms of class-average images using the LPQ technique as mentioned above. The KL distance between the histogram of the probe image sub-block and the corresponding histogram in the class-average image associated with the class label, which is attributed to that sub-block by the local SC classifier, is calculated. This value yields a criterion for the corrupted information content in the image sub-block due to illumination variation artefacts or partial occlusion. The higher value of distance concludes to the lower KL-weight (KLW) of the associated sub-region in order to reduce the influence of its local decision on the final identity decision of the probe image,

$$KLW(r,s) = 1 - \sum_{j=1}^{m} f_j^{p^{b_{rs}}} \times \log\left(\frac{f_j^{p^{b_{rs}}}}{f_j^{\bar{g}_k{}^{b_{rs}}}}\right), r,s = 1,2,\dots,\frac{M}{w}, \tag{8.31}$$

where $f_j^{p^{b_{rs}}}$ and $f_j^{\bar{g}_k{}^{b_{rs}}}$ are the j^{th} histogram bin in the b_{rs} sub-block of the probe image and class-average image, respectively. The class-average image for each class, $\bar{g}_k$, is calculated using the following averaging equation,

$$\bar{g}_k = \frac{1}{N_k}\sum_{l=1}^{N_k} g_{k,l}\ , \qquad k = 1, 2, \dots, N_c, \tag{8.32}$$

where $g_{k,l}$ is the l^{th} gallery sample of k^{th} class and N_k is the number of gallery images belonging to that class.

8.2.2.5 Majority voting scheme

Decision-fusion of the class labels yielded by the local classifiers on the probe image sub blocks, is utilized by a straight forward voting strategy, without training requirement, where the calculated local weight values are employed to discriminate between image sub blocks regarding their discrimination power. The summation on the weighted local votes for each class label is calculated and the class argument of the maximum value of the summation is the final decision with the majority of votes as follows [52, 53],

$$V(p) = \frac{1}{(\frac{M}{w})^2}\sum_{r=1}^{\frac{M}{w}}\sum_{s=1}^{\frac{M}{w}} W(r,s) \times d(r,s), \tag{8.33}$$

$$D = argmax_k\{V(p)\}, \tag{8.34}$$

where $V(p)$ is the summation of local votes of probe image sub blocks and D is the final decision on the class label of probe image.

8.2.3 Fusion of Global and Local Recognition Results

In this section, for each probe image with unknown identity, we have two class label results yielded by global and local recognition techniques, which might be agreed or disagreed on the identity of the probe face. We need to combine these results and find the final decision. Therefore, we calculate fuzzy membership (FM) using the similarity score between the intensity

values of probe image and two class-average images. Finally, the final decision on the identity of the probe image is the class label to which the probe image has greater membership. Class-average images are the average of gallery images associated with the class labels which are assigned to the probe image by two recognition strategies. In order to decide between two class labels which are attributed to the probe image by global and block-based techniques, we compared their confidence values using the fuzzy membership to each class. Therefore, we require a dissimilarity score between the intensity value of the probe image and the prototype image of each class, using a distance metric. The class prototype which is proposed is the class-average image as described in section 8.2.2.4. Thus, for each of the two class labels, we calculate the cosine distance between the probe image and the class-average of that class label. The cosine distance between two vectors A and B is as follows,

$$cos\_dis(A,B) = \frac{\sum_{j=1}^{N} a_j \times b_j}{\sqrt{\sum_{j=1}^{N}(a_j)^2} \times \sqrt{\sum_{j=1}^{N}(b_j)^2}}, \quad where, \ A = [a_1, a_2, \dots, a_N], B = [b_1, b_2, \dots, b_N]. \quad (8.35)$$

The fuzzy class membership to each of the two class labels, yielded by global and block-based methods, is calculated using the following equation [54, 55],

$$FM_{c_j}(p) = \frac{\left[cos\_dis(p, \bar{g}_{c_j})\right]^{-\frac{2}{m-1}}}{\sum_{k=1}^{2}\left[cos\_dis(p, \bar{g}_{c_j})\right]^{-\frac{2}{m-1}}}, \quad j = 1,2, \quad (8.36)$$

where $FM_{c_j}(p)$ is the fuzzy membership value of probe image, p, to the class c_j and $\bar{g}_{c_j}$ is the class-average of that class and $m > 1$ is the fuzzy parameter, 2 in this chapter.

8.3 EXPERIMENTAL RESULTS

8.3.1 Face Databases

In order to evaluate the performance of the proposed face recognition algorithm, we employed six different face databases, Extended Yale B, AR, CMU Multi-PIE, LFW, Ferret and FRGC. The robustness of proposed method against illumination variation and partial occlusion in

uncontrolled imaging conditions is assessed and compared with the results of the state-of-the-art face recognition algorithms.

8.3.1.1 Extended Yale B databases

The Extended Yale B database contains 21888 Images of 38 subjects [56], under 9 poses and 64 illumination conditions per pose. For each class of subject, 64 images with frontal pose are employed in this chapter and divided into five subsets regarding their lighting angle, α. Subset 1 with 7 images per individual, ($\alpha < 12^{\circ}$), is used as the gallery set. Subsets 2 to 5, with ($20^{\circ} < \alpha < 25^{\circ}$), ($35^{\circ} < \alpha < 50^{\circ}$), ($60^{\circ} < \alpha < 77^{\circ}$) and ($\alpha > 78^{\circ}$), which include 12, 12, 14 and 19 images per subject, respectively, are considered as four probe sets [32]. Therefore, we have 226 images in the gallery and 454, 454, 532 and 722 faces in the probe sets, respectively. Figure 8.7a shows some eye aligned and cropped sample images of one of the individuals in the database. Recognition accuracies of four probe sets are shown in Table 8.1. The results show the effectiveness of the proposed technique in severe illumination variation. 1 and 6 misclassification in subsets 4 and 5 are due to images which are corrupted files in database.

8.3.1.2 AR database

AR database contains 3536 facial images of 136 people (76 men and 60 women) [59]. In this chapter, we have chosen a subset of 2600 images of 100 subjects (50 male and 50 female). Each individual has 26 images taken in two different sessions within 2 weeks (13 images per session). This is a more challenging database comparing to Extended Yale B regarding the number of subjects, changes caused by two-week gap in photo taking and also image degradation conditions

Table 8.1. Accuracy percentage of the face recognition algorithms for subsets 2-5 of the Extended Yale B database (%).

Method	Subset2	Subset3	Subset4	Subset5
LRC[57]	100	100	83.27	33.61
LBP-MV[34]	100	100	89.66	84.90
Il+PCA+WVS[58]	100	100	97.91	96.54
DD-DTCWT[18]	100	100	98.68	99.03
wGWF[14]	100	100	97.53	95.66
LGBFR [10]	100	100	99.81	99.17
Proposed Method	100	100	99.81	99.17

due to illumination variation (left, right and all side light), facial expression (neutral, smile, scream and anger) and partial occlusion (scarf or sun glasses) [59]. Some random images from AR database are shown in Fig 8.7b. We take 7 non-occluded images per subject from session 1 and 2 as the gallery and probe sets with 700 images per dataset, respectively, in order to evaluate the performance of the algorithm in medium lighting variation, facial expression and appearance changes within two weeks as shown in Table 8.2.

8.3.1.3 CMU Multi-PIE

Carnage Melon University multi pose, illumination and expression (CMU Multi- PIE) database contains 750,000 images of 337 subjects which were taken at four sessions under 15 pose and 19 illumination conditions with smile, squint and surprise expressions [67]. Figure 8.7c illustrates some samples of Multi-PIE database. In this chapter, the images of 249 subjects from session 1 with neutral expression, frontal pose and seven severe illumination conditions, {0, 1, 7, 13, 14, 16 and 18}, are utilized as the gallery samples. The probe set includes frontal images with smile expression and four illumination conditions, {0, 2, 7 and 13}, from session 1, in order to keep the consistency in comparison [33]. Table 8.3 shows the identification accuracy.

8.3.1.4 LFW

Labeled faces in the wild (LFW) database has 13,233 images of 5749 people, which are downloaded from the web [68]. Only 1680 of individuals have two or more samples and the rest contain one sample per person. LFW-a contains aligned versions of LFW images [69]. Images in

Table 8.2. Accuracy percentage of the face recognition algorithms on AR database (%).

Method	% Recognition Rate
WSRC[60]	94.43
D-HLDO[61]	93
DFR[62]	93.7
GRRC-L2[63]	97.3
CRC-RLS[64]	93.7
CRML[33]	95.1
RSC[65]	96
LCCR+Cityblock [66]	95.86
LBP-MV[34]	99
LGBFR [10]	99
Proposed Method	99.86

Table 8.3. Accuracy percentage of the face recognition algorithms on CMU Multi-PIE database (%).

Method	LBP-MV [34]	LGBFR [10]	SRC [32]	CRC-RLS [64]	CRML [33]	Proposed Method
Recognition Accuracy (%)	75.80	91.97	93.9	94.1	94.4	94.4

Table 8.4. Accuracy percentage of the face recognition algorithms for different subsets of LFW database (%).

% Recognition Accuracy				Number of gallery sample/class	Number of probe sample/class	Number of class
DD-DTCWT [18]	LGBFR [10]	LBP-MV [34]	Proposed Method			
71.43	63.81	63.81	71.43	5	5	21
74.54	74.54	72.73	78.19	6	5	11
72.50	73.33	70	77.5	6	6	20
67.78	83.33	50	86.67	7	6	15
80.95	87.62	85.71	91.43	7	7	15
89.43	91.43	92.86	94.28	8	7	10
76.25	95	96.25	96.25	8	8	10

the database are degraded due to illumination and pose variation, low resolution, blur,race, aging effect and partial occlusion which evaluate the face identification performance of the proposed technique in real-world scenarios. Sample images of the LFW database are shown in Fig 8.7d. In this chapter, we randomly employed subjects with 10-16 samples and less than 30 degree pose variation [18]. For each individual we use the first half of samples as gallery and the second half as probe images. If the number of samples is an odd value, we have taken the first half minus one image in the probe set and the rest in the gallery set. Table 8.4 shows recognition accuracy for each subset.

8.3.1.5 FERET database

The face recognition technology (FERET) database [70] is a huge database which contains 14,051 images of 1199 people in 15 photographing sessions from 1993 to 1997. Different subsets of the database contain images in various illumination conditions, facial expression and pose variations. In this chapter we have employed subsets of 200 individuals with single sample per subject as ba (regular frontal images), bj (frontal images corresponding to ba with alternative facial expressions) and bk (frontal images corresponding to ba with different illumination conditions) [18]. 200 images in ba are used as the probe set and 400 images of subsets bj and bk,

Table 8.5. Accuracy percentage of the face recognition algorithms on FERET database (%).

Method	DT-CWT [71]	SP [72]	LBP-MV [34]	DD-DTCWT [18]	LGBFR [10]	Proposed Method
Recognition Accuracy (%)	94	91.72	95	96	96.5	99

together, are utilized as the gallery set. Figure 8.7e shows some images of probe set. Table 8.5 illustrates assessment of performance of the proposed algorithm in term of recognition accuracy.

8.3.1.6 FRGC database

In order to evaluate the performance of proposed algorithm on a large scale database with very challenging degradation conditions such as illumination variation, blur and aging effect, low resolution, head pose and facial expression, we use the images from experiment 4 of FRGC version 2. FRGC 2.0.4 contains 12776 training faces, 16028 target faces and 8014 query faces [73]. Some sample images are shown in Fig 8.7f. We have conducted two experiments on FRGC. First, following the experimental settings in [11], we choose one random image per person from 16028 target images as the gallery set and one sample per individual from 8014 query images as the probe images. Thus, we have 466 images in either of gallery or probe sets. The average recognition accuracy is illustrated in Table 8.6.

Furthermore, in the second experiment, we have employed one sample per individual of the target set as the gallery set of 466 faces and utilize the 8014 query samples as the probe set. Table 8.7 shows the recognition result compare to the state of the art techniques [74].

8.3.2 Performance of the Algorithm under Partial Occlusion

8.3.2.1 Real occlusion

In this section we conduct three different experiments on the images from AR database which contain partial occlusion, sun glasses and scarf. Figure 8.8 shows some occluded samples in AR database. In the first test, we have utilized four non-occluded images per individual, with neutral and different facial expressions, in session 1 and 2 to generate a gallery set of 800 images. Two probe sets (200 occluded images in each set) include neutral images with sun glasses and scarf from both sessions [33, 63].

Table 8.6. Accuracy percentage of the face recognition algorithms for the first experiment on FRGC 2.0.4 database (%).

Method	LBP-MV [34]	Gabor-PCA [75]	LOG-DCT [19]	OLHE [6]	LGBFR [10]	Proposed Method
Recognition Accuracy (%)	10.51	13.73	15.82	20.53	32.62	40.77

Table 8.7. Accuracy percentage of the face recognition algorithms for the second experiment on FRGC 2.0.4 database (%).

Method	LBP-MV [34]	AGL [76]	NN-ADA-AM [74]	NN-ADA-RM [74]	LGBFR [10]	Proposed Method
Recognition Accuracy (%)	6.03	15.49	18	18	23.3	32.43

Figure 8.7. Sample images of different databases: (a) Extended Yale B, (b) AR, (c) CMU Multi-PIE, (d) LFW, (e) FERET and (f) FRGC 2.0.4.

In the second experiment, we have taken fourteen non-occluded images per subject from session 1 and 2 as the gallery set of 1400 images and 6 occluded images per individual with scarf and 6 with sun glasses from both sessions as two probe sets of 600 images each, respectively [66]. Then, in the third challenge, we have repeated the second experiment with a smaller gallery set of 200 images, by employing just two non-occluded images with neutral expression in session 1 and 2 [39, 77]. Recognition results of three mentioned experiments are compared with the state of the art techniques in Table 8.8, 8.9 and 8.10, respectively.

Figure 8.8. Samples of occluded images in AR database from both photography sessions.

Table 8.8. Accuracy percentage of the face recognition algorithms for the first experiment on the occluded AR database (%).

Method		SRC [32]	RPCA [78]	CRML [33]	GRRC-L2 [63]	RSC [65]	LGBFR [10]	LBP-MV [34]	Proposed Method
Recognition Rate (%)	Sun	87	90.5	93	93	99	100	100	100
	Scarf	59.5	89.5	90.5	79	97	99.5	99.5	100

8.3.2.2 Block occlusion

In this experiment we have aimed at evaluating the behaviour of the proposed algorithm in the presence of partial occlusion on the images which are highly affected by illumination variation. In order to do this experiment we have located a window of mandrill image as shown in Fig 8.9a, randomly on the probe image and varied the size of occlusion block to find the point at which the proposed algorithm breaks down (recognition rate reduces to less than 90% [79]). In order to keep the consistency in comparison, the combination of subsets 1 and 2 of the Extended Yale B database, 722 images, is chosen as the gallery set and subset 3, 456 images, is taken as the probe set. The size of occlusion block is changed from 0% to 50% of the probe image area [63]. Figure 8.10 illustrates recognition rate versus the percentage of the occlusion coverage and show the superior performance of the proposed method in comparison with other techniques. In all cases we have 100% recognition rate. We have repeated the experiment with a black box, instead of mandrill image, as shown in Fig 9b, which does not contain any information. Figure 8.11 shows the recognition accuracy, compared to the mandrill occlusion with size of occlusion coverage from 0% to 90% of the probe image area. The break down point of the proposed algorithm occurs at 90% for mandrill occlusion while for black box it occurs at 80% coverage. This experiment shows that even with 80% mandrill occlusion and 70% black block occlusion, the algorithm still can recognize faces at 90.57% and 97.37% accuracy percentages.

Table 8.9. Accuracy percentage of the face recognition algorithms for the second experiment on the occluded AR database (%).

Method		LRC [57]	CRC/RLS [64]	SRC [32]	LCCR [66]	CESR [80]	LGBFR [10]	LBP-MV [34]	Proposed Method
Recognition Rate (%)	Sun	88.17	88.33	93	91.17	97.50	100	100	100
	Scarf	91.38	92.17	91.83	92.50	91.33	99.17	99.5	99.83

Table 8.10. Accuracy percentage of the face recognition algorithms for the third experiment on the occluded AR database (%).

Method	Sunglasses	Scarf
E-GV-LBP-M[77]	47.22	82.78
E-GV-LBP-P[77]	44.07	86.67
SADT[39]	95.5	75
LBP-MV[34]	98.5	96.5
LGBFR [10]	99	95.5
Proposed Method	99.67	97.17

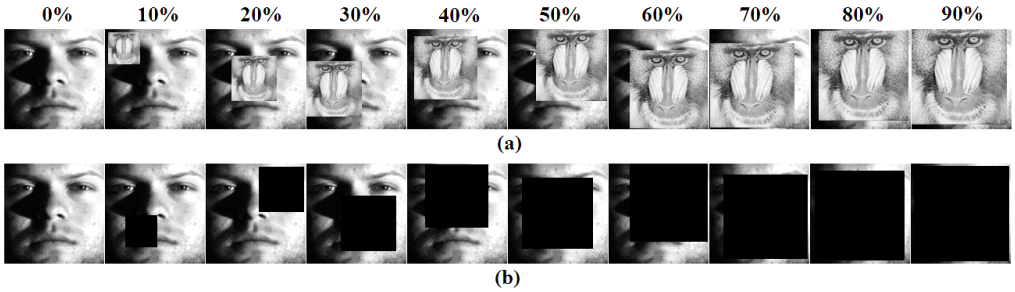

Figure 8.9. Sample images of Extended Yale B subset 2 with occlusion coverage from 0% to 90% of the face area: (a) mandrill block as occlusion and (b) black block as occlusion.

8.4 CONCLUSION

In this chapter, a modified technique is presented for human face recognition, by combining the global and local based approaches to achieve improved recognition rate for images with degraded conditions. A preprocessing technique based on fusion of multi-scale Weber approach and DT-DD-CWT enhancement, yields the illumination insensitive representation of image. 2D-Gabor filter followed by PCA is applied on the normalized face image to extract the global

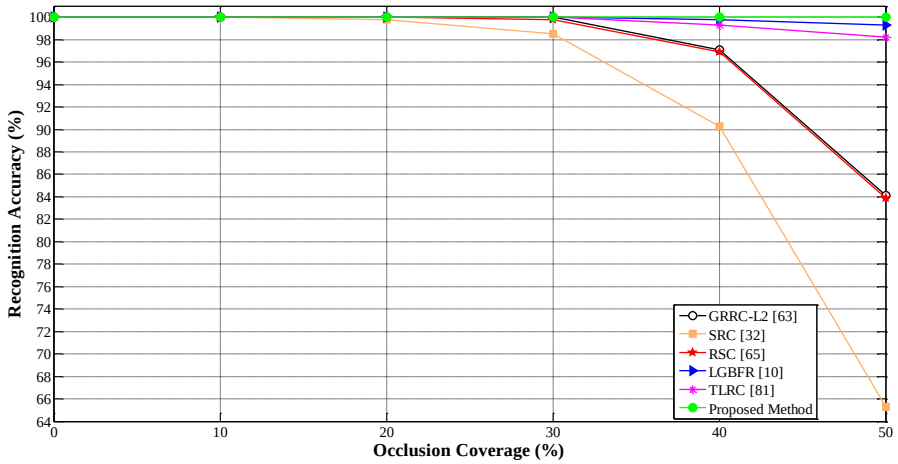

Figure 8.10. Recognition accuracy (%) of different techniques versus the percentage of occlusion (mandrill block) coverage of the face area.

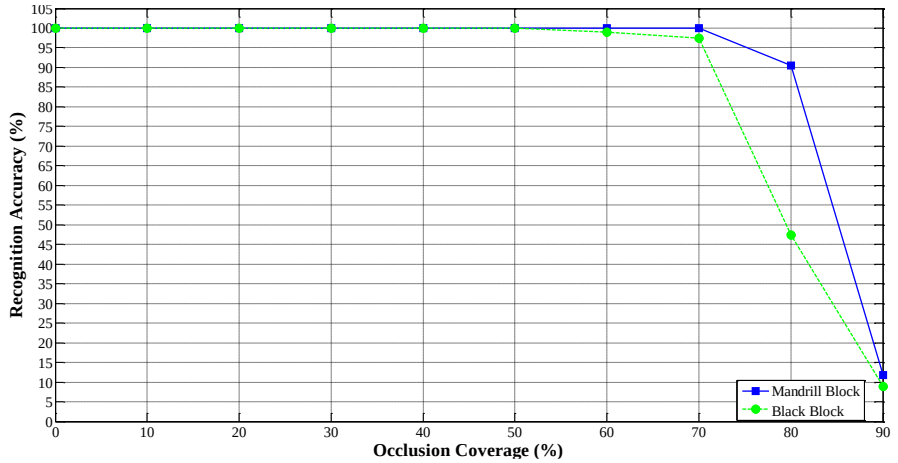

Figure 8.11. Recognition accuracy (%) of the proposed algorithm versus the percentage of occlusion (mandrill and black block) coverage of the face area.

feature vector of image which is classified by an ELM technique. Simultaneously, the preprocessed image passes through a local texture descriptor (LPQ) and the result is partitioned into sub blocks. Local histograms are classified using local sparse classifiers and their decisions are combined in a majority voting decision fusion. Finally, the global-based and block-based class label results are fed into the fuzzy membership calculation and the class label with larger membership value is selected as the final identity. The proposed global and local based

approaches provide complementary information of shape and texture of the face image. Knowing that Gabor is an effective approach to extract a large number of distinctive features and ELM is a classifier with significant discrimination power and robustness against degraded conditions. Unfortunately, global-based technique will not perform well when occlusions exist. Therefore, its fusion with a block-based approach containing a powerful local descriptor, LPQ, and an occlusion robust classifier, SC, improves the recognition accuracy. Effectiveness of the proposed algorithm is evaluated on Extended Yale B, AR, CMU Multi-PIE, LFW, FERET and FRGC databases with images under different degraded conditions and large number of images. The experimental results show the significantly improved performance of the proposed algorithm under severe illumination, partial occlusion and low quality images in uncontrolled imaging conditions. Our best recognition rate using FRGC 2.0.4 indicated 28.15% improvement over the state of the art techniques in this area.

8.5 REFERENCES

[1] T. Charoenpong, "A survey of face occlusion detection for visual surveillance system," in *Int. Conf. Knowl. Smart Tech. (KST10)*, 2010, pp. 24-25.

[2] H. Han, S. Shan, X. Chen, W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, (6) pp. 1691-1699, 2013.

[3] J. R. Solar, J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different preprocessing approaches," *Pattern Recognition Letters*, vol. 29, (14) pp. 1966-1979, 2008.

[4] S. M. Pizer, E. P. Amburn, J. D. Austin, et.al., "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Processing, vol. 39, (3) pp. 355-368, 1987.

[5] S. Shan, W. Gao, B. Cao, D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *IEEE. Int. Wksp. Anal. Mod. (AMFG03)*, Nice, 2003, pp. 157-164.

[6] P. H. Lee, S. W. Wu, Y. P. Hung, "Illumination compensation using oriented local histogram equalization and its application to face recognition," *IEEE Transactions on Image Processing*, vol. 21, (9) pp. 4280-4289, 2012.

[7] H. Wang, S. Z. Li, Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *IEEE Int. Conf. Autom. FGR. (AFGR04)*, Seoul, 2004, pp. 819-824.

[8] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing*, vol. 18, (11) pp. 2599-2606, 2009.

[9] B. Xu, T. P. Zhang, Z. W. Shang, "Multi-scale invariant abstracted under varying illumination," in *IEEE Int. Conf. Wavelet. Anal. Pattern Recogn. (ICWAPR12)*, Xian, 2012, pp. 28-32.

[10] S. Nikan, M. Ahmadi, "A Local Gradient-Based Illumination Invariant Face Recognition using LPQ and Multi-Resolution LBP Fusion," *IET Image Processing*, 10 pp, 2014, in press.

[11] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, "Subspace learning from image gradient orientations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, (12) pp. 2454-2466, 2012.

[12] A. K. Jain, *Fundamentals of Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall.

[13] B. Wang, W. Li, W. Yang, Q. Liao, "Illumination normalization based on Weber's law with application to face recognition," *IEEE Signal Processing Letters*, vol. 18, (8) pp. 462-465, 2011.

[14] Y. Wu, Y. Jiang, Y. Zhou, *et al*., "Generalized Weber-face for illumination-robust face recognition," *Neurocomputing*, vol. 136, pp. 262-267, 2014.

[15] S. Du, R. Ward, "Wavelet-based illumination normalization for face recognition," in *IEEE Int. Conf. Image Process. (ICIP05)*, Genova, 2005, pp. II - 954-7.

[16] M. Emadi, M. Khalid, R. Yusof, F. Navabifar, "Illumination normalization using 2D wavelet," *Procedia Engineering*, vol. 41, pp. 854-859, 2012.

[17] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, (3) pp. 234-253, 2001.

[18] A. Baradarani, Q. M. J. Wu, M. Ahmadi, "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering," *Pattern Recognition*, vol. 46, (1) , pp. 57-72, 2013.

[19] W. Chen, M. J. Er, S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 36, (2) pp. 458-466, 2006.

[20] T. Ahonen, A. Hadid, M. Pietikainen, "Face recognition with local binary patterns," *Springer Computer Vision*, vol. 3021, pp. 469-481, 2004.

[21] D. Zang, "Illumination invariant object tracking based on multiscale phase," in *IEEE International Conference on Image Processing (ICIP10)*, Hong Kong, 2010, pp. 357-360.

[22] J. B. Li, S. C. Chu, J. S. Pan, "Statistical learning-based face recognition," in *Business Media*, New York, 2014, pp. 19-47.

[23] J. Daugman, "Complete discrete 2-d Gabor transform by neural networks for image analysis and compression," *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 36, (7) pp. 1169-1179, 1988.

[24] X. Tan, B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, (6) pp. 1635-1650, 2010.

[25] T. Y. Kim, K. M. Lee, S. U. Lee, C. H. Yim, "Occlusion Invariant Face Recognition Using Two-Dimensional PCA," *Communications in Computer and Information Science*, vol. 4, pp. 305-315, 2007.

[26] Z. C. Lian, M. J. Er, J. Li, "A novel face recognition approach under illumination variations based on local binary pattern," in *Int. Conf. Comp. Anal. (CAIP11)*, Seville, 2011, pp. 89-96.

[27] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, (6) pp. 748-763, 2002.

[28] H. J. Oh, K. M. Lee, S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images," *Image and Vision Computing*, vol. 26, (11) pp. 1515-1523, 2008.

[29] H. R. Kanan, K. Faez, Y. Gao, "Face recognition using adaptively weighted patch PZM array from a single exemplar image per person," *Pattern Recognition*, vol. 41, (12) pp. 3799-3812, 2008.

[30] H. R. Kanan, K. Faez, "Recognizing faces using adaptively weighted sub-Gabor array from a single sample per enrolled subject," *Pattern Recognition*, vol. 28, (3) pp. 438-448, 2010.

[31] H. K. Ekenel, R. Stiefelhagen, "Why is facial occlusion a challenging problem?", in *Int. Conf. Advances in Biometrics (ICB09)*, Alghero, 2009, pp. 229-308.

[32] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, (2) pp. 210-227, 2009.

[33] M. Yang, Z. Feng, S. Shiu, L. Zhang, "Fast and robust face recognition via coding residual map learning based adaptive masking," *Pattern Recognition*, vol. 47, (2) pp. 535-543, 2014.

[34] S. Nikan, M. Ahmadi, "Human face recognition under occlusion using LBP and entropy weighted voting," in *International Conference on Pattern Recognition (ICPR12)*, Tsukuba, November 2012, pp. 1699-1702.

[35] W. Zhang, S. Shan, X. Chen, W. Gao, "Local Gabor binary patterns based on Kullback–Leibler divergence for partially occluded face recognition," *IEEE Signal Processing Letters*, vol. 14, (11) pp. 875-878, 2007.

[36] Y. Su, S. Shan, X. Chen, W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Transactions on Image Processing*, vol. 18 (8) pp. 1885-1896, 2009.

[37] R. Chellappa, C. Wilson, S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 85, (5) pp. 705-741, 1995.

[38] Y. Yan, H. Wang, D. Suter, "Multi-subregion based correlation filter bank for robust face recognition," *Pattern Recognition*, 2014, in press.

[39] R. Mehta, J. Yuan, K. Egiazarian, "Face recognition using scale-adaptive directional and textural features," *Pattern Recognition*, vol. 47, (5) pp. 1846-1858, 2014.

[40] I. W Selesnick, "The double-density dual-tree DWT," *IEEE Transactions on Signal Processing*, vol. 52, (5) pp. 1304-1314, 2004.

[41] K. H. Pong, K. M. Lam, "Multi-resolution feature fusion for face recognition," *Pattern Recognition*, vol. 47, (2) pp. 556-567, 2014.

[42] L. Zhang, D. Tjondronegoro, V. Chandran, "Random Gabor based templates for facial expression recognition in images with facial occlusion," *Neurocomputing*, 2014, in press.

[43] Q. Zhang, H. Hang, C. Ji, *et al*., "Gabor-based anisotropic diffusion for speckle noise reduction in medical ultrasonography," *Journal of the Optical Society of America A*., vol. 31, (6) pp. 1273-1283, 2014.

[44] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *IEEE International Conference on Computer Vision (ICCV05)*, Beijing, 2005, pp. 786-791.

[45] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, (1-3) pp. 489-501, 2006.

[46] C. H. Chan, M. A. Tahir, J. Kittler, M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, (5) pp. 1164-1177, 2013.

[47] Y. Li, S. Shan, H. Zhang, S. Lao, X. Chen, "Fusing magnitude and phase features for robust face recognition," in *Asian Conference on Computer Vision (ACCV12)*, Daejeon, 2012, pp. 601-612.

[48] S. Jayaraman, S. Esakkirajan, T. Veerakumar, *Digital Image Processing*, New Delhi: Tata McGraw Hill, 2009.

[49] V. Ojansivu, J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *International Conference on Image and Signal Processing (ICISP08)*, Cherbourg-Octeville, 2008, pp. 236-243.

[50] M. R. Banham, A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, (2) pp. 24-41, 1997.

[51] W. Ou, X. You, D. Tao, *et al*., "Robust face recognition via occlusion dictionary learning," *Pattern Recognition*, vol. 47, (4) pp. 1559-1572, 2014.

[52] N. G. Kasapoğlu, T. Eltoft, "Decision fusion of classifiers for multifrequency PolSAR and optical data classification," in *IEEE International Conference on Recent Advances in Space Technologies (RAST13)*, Istanbul, 2013, pp. 411-416.

[53] H. Yang, Q. Du, B. Ma, "Weighted decision fusion for supervised and unsupervised hyperspectral image classification," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS10)*, Honolulu, 2010, pp. 3656-3659.

[54] J. M. Keller, M. R. Gray, J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, (4) pp. 580-585, 1985.

[55] J. Derrac, S. Gracia, F. Herrera, "Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects," *Information Sciences*, vol. 260, pp. 98-119, 2014.

[56] K. C. Lee, J. Ho, D. Kriegman, "Acquiring linear sub spaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, (5) pp. 684-698, 2005. The Extended YaleB face database, Available (online) at:
<http://vision.ucsd.edu /~leekc/ExtYaleDatabase/ExtYaleB.html >.

[57] I. Naseem, R. Togneri, M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, (11) pp. 2106-2112, 2010.

[58] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, "Illumination invariant feature extraction and mutual information based local matching for face recognition under illumination variation and occlusion," *Pattern Recognition*, vol. 44, (10-11) pp. 2576-2587, 2011.

[59] A. Martinez, R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998. Available (online) at: < http:// www2.ece.ohio-state.edu/~aliex/ARdatabase.html >.

[60] C. Y. Lu, H. Min, J. Gui, L. Zhu, Y. K. Lei, "Face recognition via weighted sparse representation," *Visual Communication and Image Representation*, vol. 24, (2) pp. 111-116, 2013.

[61] J. Qian, J. Yang, G. Gao, "Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction," *Pattern Recognition*, vol. 46, (10) pp. 2724-2739, 2013.

[62] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Transactions on Information Forensics and Security*, vol. 7, (3) pp. 954-965, 2012.

[63] M. Yang, L. Zhang, S. C. K. Shiu, D. Zhang, "Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary," *Pattern Recognition*, vol. 46, (7) 1865-1878, 2013.

[64] L. Zhang, M. Yang, X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *IEEE International Conference on Computer Vision (ICCV11)*, Barcelona, 2011, pp. 471-478.

[65] M. Yang, L. Zhang, J. Yang, D. Zhang, "Robust sparse coding for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR11)*, Providence, 2011, pp. 625-632.

[66] X. Peng, L. Zhang, Z. Yi, K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognition*, vol. 47, (9) pp. 2794-2806, 2014.

[67] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, (5) pp. 807-813, 2010.

[68] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," *Technical Report University of Massachusetts*, 2007.

[69] L. Wolf, T. Hassner, Y. Taigman, "Effective face recognition by combining multiple descriptors and learned background statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, (10) pp. 1978-1990, 2011.

[70] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, (10) pp. 1090-1104, 2000.

[71] A. Eleyan, H. O¨zkaramanli, H. Demirel, "Complex wavelet transform-based face recognition," *EURASIP Journal on Advanced Signal Processing*, (195) pp. 1-13, 2008.

[72] M. El Aroussi, M. El Hassouni, S. Ghouzali, M. Rziza, D. Aboutajdine, "Local appearance based face recognition method using block based steerable pyramid transform," *Signal Processing*, vol. 91, (1) pp. 38-50, 2011.

[73] P. J. Phillips, P. Flynn, T. Scruggs, *et al*, "Overview of the face recognition grand challenge," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, San Diego, 2005, pp. 947-954.

[74] M. Kana, S. Shana, Y. Suc, D. Xub, X. Chena, "Adaptive discriminant learning for face recognition," *Pattern Recognition*, vol. 46, (9) pp. 2497-2509, 2013.

[75] L. Chengjun, H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Transactions on Neural Networks*, vol. 14, (4) pp. 919-928, 2003.

[76] Y. Su, S. Shan, X. Chen, W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR10)*, San Francisco, 2010, pp. 3699-2706.

[77] Z. Lei, S. Liao, M. Pietikäinen, S. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Transactions on Image Processing*, vol. 20, (1) pp. 247-256, 2011.

[78] X. Luan, B. Fang, L. Liu, W. Yang, J. Qian, "Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion," *Pattern Recognition*, vol. 47, (2) pp. 495-508, 2014.

[79] X. X. Li, D. Q. Dai, X. F. Zhang, C. X. Ren, "Structured sparse error coding for face recognition with occlusion," *IEEE Transactions on Image Processing*, vol. 22, (5) pp. 1889-1900, 2013.

[80] R. He, W. S. Zheng, B. G. Hu, "Maximum correntropy criterion for robust face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, (8) pp. 1561-1576, 2011.

[81] J. Lai, X. Jiang, "Robust face recognition using trimmed linear regression," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP13)*, Vancouver, 2013, pp. 2979-2983.

Chapter 9

Conclusion and Future Work

9.1 SUMMARY AND CONCLUSION

Illumination variation affects the accuracy of face recognition techniques significantly. An illumination insensitive preprocessing technique based on the image gradient was proposed in this dissertation. Image representation using the ratio between gradient magnitude and the current intensity level of image was shown to be insensitive against severe levels of lighting effect. Also, the proposed combination of multi-scale Weber analysis and enhanced DD_DT_CWT was demonstrated to have an improved performance when illumination variation is present. The ratio of differences between the gray value of each pixel and its neighbors at different scales and the current intensity level of pixel leads to an illumination invariant representation. This was further improved by decomposing the Weber analyzed image into its frequency components using complex wavelet transform as a powerful shift invariant and directionally selective function with contrast enhancement of low frequency components and enlarging the intensity level of detail subbands. The proposed techniques achieved improved recognition rates for images which are highly affected by illumination variation. Furthermore, the proposed utilization of the illumination insensitive image descriptors, such as LBP, LPQ and Gabor filter, on the preprocessed images leads to further robustness against illumination effect.

Applying lower resolution images reduces the computational complexity and memory requirement. The proposed evaluations on the performance of feature extraction and classification techniques when images have low-resolution, showed that they are affected differently by resolution reduction. Gabor technique is a powerful feature extractor to provide a huge number of distinctive features and leads to acceptable recognition performance. However, its recognition accuracy is reduced by decreasing the image resolution. While, the performance of FLD is improved by reducing the image size due to the fact that within class differences are reduced in lower resolutions thus discrimination power is increased. However, DCT and PCA showed considerable robustness against resolution changes. Analyzing the low-frequency information using LPQ, shows more robustness against changes in the image size. Resolution enhancement using the proposed interpolation techniques improves the recognition accuracy and nearest neighbor technique outperforms bilinear and bicubic interpolation methods. In the proposed evaluation on the performance of classification strategies, ELM and MLFNN, and SC among different classification techniques, show better performance under degraded conditions.

In real-world applications, some parts of discriminative facial information are covered by partial occlusion which reduces the accuracy of recognizing people significantly. In the proposed block-based approach the effect of occluded parts of image on the identification process is decreased by devoting weights to the image subblocks according to the importance of their information properties. Local entropy, class posterior probability and Kullback–Leibler divergence are proposed as the local weights in the score or decision level fusion to decide on the class label of the unknown individual. The proposed strategies lead to noticeable improvement in the recognition accuracies compared to the state-of-the-art algorithms on the artificially occluded images or face databases with real occlusion such as sunglasses or scarf.

According to the psychological point of view about the human perception in face recognition, and employing the evaluation results in the previous chapters, the proposed hierarchical structure in the final part of this thesis employed the fusion of global and local information to improve the recognition robustness against different conditions of image degradation. In the global technique, Gabor filter is applied on the whole face image to extract a large number of features with significant discrimination power. PCA is employed as a dimensionality reduction technique which is easy to execute and robust against resolution reduction. ELM as a powerful classifier is

applied to attribute the global class label to the unknown individual. Furthermore, through the local-based approach, illumination insensitive LPQ descriptor is applied on the preprocessed image where the low frequency local texture information is extracted. Subsequently, the resultant image is divided into sub blocks and local histograms are fed into the sparse classifiers. Histogram based features are robust against rotation and scale variation. Gabor and LPQ provide complementary shape-texture and magnitude-phase information. Complementary performance of global and local techniques against various degraded conditions leads to considerable improvement in the face recognition accuracy. Effectiveness of the proposed algorithm is evaluated on Extended Yale B, AR, CMU Multi-PIE, LFW, FERET and FRGC databases with images under different degradation conditions and large number of images. The experimental results show the significantly improved performance of the proposed algorithm under severe illumination, partial occlusion and low quality images in an uncontrolled imaging conditions. Our best recognition rate using FRGC 2.0.4 which is the most challenging face database with images under uncontrolled conditions indicated 28.15% improvement over the state-of-the-art techniques in this area. Also, experimental evaluation of the proposed strategy shows that even with 70% face coverage by a black box occlusion with no information content, the algorithm still recognizes faces at 90.57% accuracy. The original contributions of this work are as follows:

- A novel illumination invariant representation has been proposed based on the ratio of gradient magnitude and current intensity of image which is easy to execute and improves the performance of the face recognition under poor illumination.
- A combination of multi-scale Weber and enhanced DD-DT-CWT was proposed as the illumination invariant preprocessing with significant performance improvement under severe lighting changes. This technique not only compensates the pixel offsets, but also enhances the contrast of image details at different scales and spatial frequencies.
- Block based face recognition was proposed in order to reduce the effect of corrupted parts of the image due to occlusion or illumination artefacts. It should be noted that since we are applying the same recognition procedures on each image sub block, this technique is a good candidate for implementation on a parallel processing hardware such as graphics processing unit (GPU), where the computational complexity can be reduced

from image scale to a small sub block scale which is significantly beneficial for real time applications.

- Decision fusion of global and local based strategies and combination of powerful and discriminative feature extractors and classification approaches, which are robust against appearance changes and explore local characteristics of facial images, has been utilized to profit from their complementary performance when different degrading conditions occur at the same time.

9.2 FUTURE WORK

The proposed block based analysis of face image improves the identification robustness against partial occlusion, illumination effect artifacts and facial expression. However, due to applying the same feature extraction and classification procedures repeatedly on the image partitions separately, we can substantially reduce the computational cost by executing the parallel processing approaches on the GPU. Therefore, the computational complexity would be reduced.

Optimizing MATLAB implementation codes of the proposed techniques in the dissertation are suggested to be translated into Verilog or VHDL hardware description languages to be realized and tested on the FPGA boards to be used in real-time applications.

Moreover, in order to deal with the head pose and illumination variation we suggest generating 3D model of face using 2D images. As shown in the proposed methods in this dissertation, fusion of complementary information can improve the recognition accuracy significantly. Therefore combination of shape, depth, texture, phase, magnitude, dense and sparse information is suggested.

VITA AUCTORIS

Soodeh Nikan received the B.Sc. and M.Sc. degrees in electrical engineering in 2004 and 2009 in Iran. She is currently pursuing her Ph.D. degree in electrical engineering with University of Windsor, Windsor, ON, Canada and is graduating in 2014. Pattern recognition, image processing, computer vision, machine learning, biomedical image analysis and telecommunication are her main research interests.